

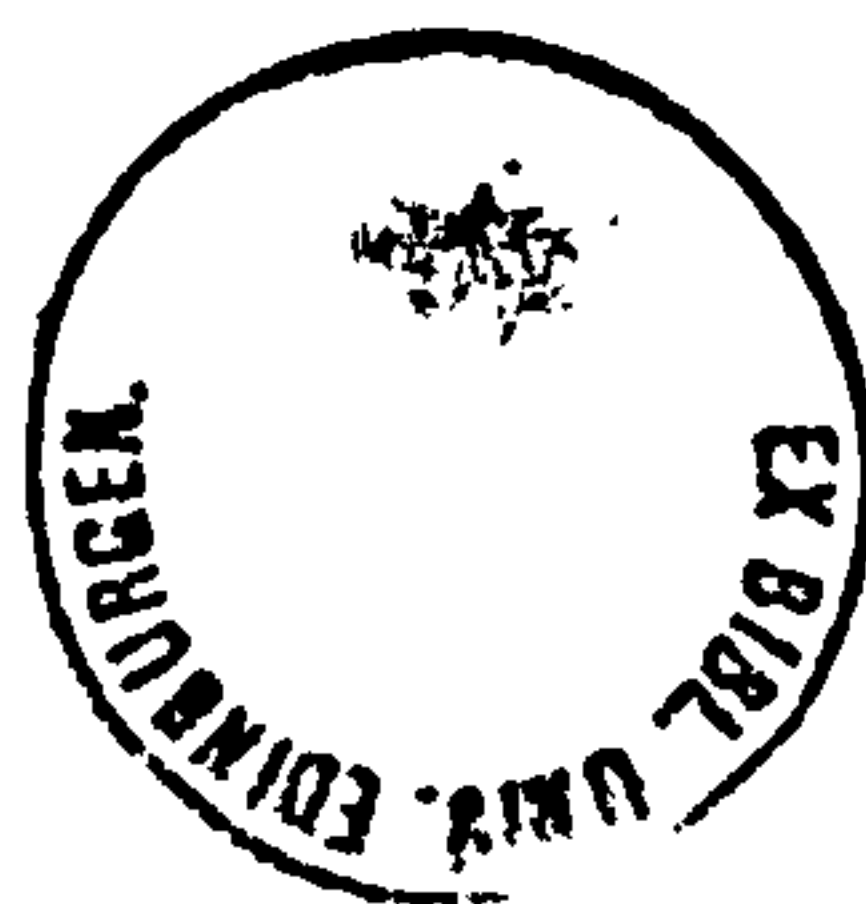
# **MECHANISMS OF VOWEL DEVOICING IN JAPANESE**

A thesis submitted in fulfilment of the requirements for  
the degree Doctor of Philosophy  
to the Faculty of Arts, the University of Edinburgh

by

Mariko Kondo

1997



## ACKNOWLEDGEMENTS

First of all I would like to thank my supervisors, Jim Hurford, Alan Kemp, Geoff Lindsey, John Laver and Ron Asher for advice, comments, criticisms and encouragement during the course of this thesis. Special thanks to Jim Hurford and Alan Kemp for their patience in reading various rough drafts of the thesis and giving me detailed and helpful comments. I am extremely grateful because Jim Hurford volunteered to be my supervisor after other supervisors had left the department and Alan Kemp spent his precious time after retirement reading my thesis.

I also would like to thank Jim Scobbie and Elizabeth Uldall for their useful comments, Ellen Bard for statistical advice, Steve Isard for technical advice, and Mike Morony for writing a computer programme which saved me hundreds of hours.

I owe thanks to the computer staff of the Linguistic Department, Morag Brown, Norman Dryden, Irene Macleod, Cedric MacMartin, Len Scott and Stewart Smith, not only for technical assistance but also a wonderful time in the computer lab, otherwise life there would have been extremely painful. And a big thanks to Ethel Jack whose encouragement and cajoling were most helpful.

In addition to people in Edinburgh, various people gave me a lot of help during my Ph.D.. I would like to express thanks to Mr. Kikuo Maekawa of the National Language Research Institute in Tokyo for letting me use their facility and a recording studio to do hours of recording, otherwise this thesis had never been completed. I am also most grateful to Professor Mary Beckman of the Ohio State University, Dr. Yoshinori Sagisaka and Dr. Nick Campbell of ATR in Kyoto, Dr. Yasuko Nagano-Madsen of Lund University, Professor Katsumasa Shimizu of Nagoya Gakuin University and Ms. Ayako Tsuchida of Cornell University. They sent me various papers which otherwise I would not have been able to get access to, and gave me useful comments and discussion on parts of my thesis.

I also thank Etsuko Hayashi, Chizuko Magata, Teruo Magata, Harumi Makino, Norimasa Morita, Mayumi Muraoka, Maho Mochizuki, Ken Oishi, Mihoko Toba and Kazuko Ueda for their patience for sitting in a recording studio, talking to a microphone for hours of recording sessions.

Finally, I would like to thank to all colleagues (too many to mention individually) doing or having completed Ph.D. for giving me a wonderful time in Edinburgh despite weather and mental torture of doing Ph.D's.

August, 1996

## **Declaration**

I declare that this thesis has been composed by myself, and all works in this theses have been planned and conducted by myself.

**Mariko Kondo**

**August, 1996**



# TABLE OF CONTENTS

	page
ACKNOWLEDGEMENTS.....	i
DECLARATION.....	iii
ABSTRACT.....	viii
 <b>CHAPTER 1 INTRODUCTION</b>	 <b>1</b>
1.1 Aims of the thesis.....	1
1.2 Overview.....	2
1.3 Outline of the thesis .....	8
 <b>CHAPTER 2 CROSS-LINGUISTIC STUDY ON THE PROCESS OF VOWEL DEVOICING</b>	 <b>11</b>
2.1 Introduction.....	11
2.2 The definition of vowel devoicing.....	12
2.2.1 Vowel devoicing, reduction, deletion and weakening.....	12
2.2.2 Phonological processes of vowel devoicing.....	13
2.2.3 Phonetic description of voiceless vowels.....	17
2.2.4 Are voiceless vowels distinctive?.....	21
2.3 Factors affecting vowel reduction, devoicing and deletion.....	23
2.3.1 Intrinsic features of vowels.....	23
2.3.2 Extrinsic factors.....	26
2.3.2.1 Phonological environments.....	26
2.3.2.2 Stress and accent.....	27
2.3.2.3 Speaking tempo and style.....	31
2.3.2.4 Sociolinguistic factors.....	32
2.3.2.5 Lexical, syntactic and semantic constraints.....	33
2.4 Processes of vowel reduction.....	35
2.4.1 Greek vowel reduction processes.....	35
2.4.2 Hebrew vowel reduction processes.....	37
2.4.3 German vowel reduction processes.....	40
2.4.4 Gestural overlap analysis of vowel devoicing in Korean.....	42
2.5 Conclusions.....	44
 <b>CHAPTER 3 JAPANESE PHONOLOGY</b>	 <b>46</b>
3.1 Introduction.....	46
3.2 Japanese syllable structure and mora.....	46
3.3 Accent and intonation.....	53
3.4 Vowels.....	55
3.5 Consonants.....	57
3.5.1 Phonemes and allophones .....	57
3.5.2 Moraic nasals.....	59
3.5.3 Moraic obstruents .....	59

<b>CHAPTER 4</b>	<b>REVIEW OF PAST WORK ON JAPANESE VOWEL DEVOICING</b>	<b>61</b>
4.1	Introduction.....	61
4.2	Phonological and phonetic definitions of voiceless vowels.....	62
4.3	The physiological mechanism of voiceless vowel production.....	65
4.4	The sociolinguistic function of vowel devoicing.....	68
4.5	Factors affecting vowel devoicing in Japanese.....	70
4.5.1	Phonetic environment.....	72
4.5.1.1	The vowel itself.....	72
4.5.1.2	The vowel in a following mora.....	74
4.5.1.3	The preceding consonant.....	75
4.5.1.4	The following consonant.....	76
4.5.2	Pitch and accent.....	79
4.5.3	Position in an utterance.....	90
4.5.4	Word boundary.....	92
4.5.5	Length of an utterance.....	93
4.5.6	Speech style.....	94
4.5.7	Speech rate.....	94
4.5.8	Devoicing in consecutive morae.....	95
4.6	Conclusions.....	98
<b>CHAPTER 5</b>	<b>THE EFFECT OF BLOCKING FACTORS IN VOWEL DEVOICING</b>	<b>100</b>
5.1	Introduction.....	100
5.2	Experimental hypotheses.....	102
5.3	Experimental methods.....	105
5.3.1	Subjects and procedure.....	105
5.3.2	Materials.....	105
5.3.2.1	Test words.....	105
5.3.2.2	Blocking factors.....	109
5.3.2.2.1	Accentuation.....	109
5.3.2.2.2	Internal word boundary.....	109
5.3.2.2.3	Effect of combined blocking factors.....	110
5.3.2.2.4	Position in an utterance.....	111
5.3.2.3	Phonetic environments for the test words and conditions.....	114
5.3.3	The criteria for determining voicing.....	118
5.4	Results and discussion.....	118
5.4.1	Problems regarding the definition of voiceless vowels.....	118
5.4.1.1	The definition of voiced and voiceless vowels.....	118
5.4.1.2	Partially voiced vowels.....	122
5.4.1.3	Vowel deletion.....	124
5.4.1.4	Phonetic realisation of voiceless vowels.....	126
5.4.1.5	Accent placement.....	129
5.4.2	Difference in individual speakers.....	129
5.4.3	Non-high vowel devoicing and high vowel devoicing in non-devoicing environments.....	131
5.4.4	Adjacent consonants.....	132
5.4.5	Position in the utterance.....	137
5.4.6	Devoicing conditions.....	138
5.4.7	The number of blocking factors and devoicing rates.....	141
5.5	Conclusions.....	145

<b>CHAPTER 6</b>	<b>THE EFFECT OF DIFFERENT SPEAKING RATES ON VOWEL DEVOICING</b>	<b>148</b>
6.1	Introduction.....	148
6.2	Experimental hypotheses.....	156
6.3	Experimental methods.....	158
6.3.1	Subjects and procedure.....	158
6.3.2	Materials.....	158
6.3.3	Speech rate setting.....	158
6.3.4	Data analysis.....	159
6.4	Experimental results and discussion.....	159
6.4.1	Effect of the type of preceding consonant.....	159
6.4.2	Effect of position in an utterance.....	161
6.4.3	Effect of devoicing condition.....	165
6.4.4	Difference in speaking style.....	173
6.4.5	High vowel devoicing in non-devoicing environments.....	177
6.4.6	Devoicing of non-high vowels.....	179
6.4.7	Problems of controlling speaking rate.....	183
6.4.8	Is high vowel devoicing the process of optional fast speech?.....	184
6.5	Conclusions.....	191
<b>CHAPTER 7</b>	<b>THE DURATION OF MORAE WITH A DEVOICED VOWEL</b>	<b>193</b>
7.1	Introduction.....	193
7.2	Experimental hypotheses.....	205
7.3	Experimental methods.....	209
7.3.1	Subjects and materials.....	209
7.3.2	Procedure.....	210
7.3.3	Durational measurements and segmentation criteria .....	211
7.4	Results and discussion.....	212
7.4.1	Durational ratio between moraic consonant and CV mora.....	212
7.4.2	Analysis by T-test.....	215
7.4.3	Problems of segmentation.....	217
7.4.4	The duration of voiced high vowels in devoicing environments.....	218
7.4.5	Durational relationship between a consonant and a vowel in the same mora.....	221
7.4.6	The duration of devoiced morae and the number of devoiced morae in a word.....	224
7.4.7	Reality of mora-timing.....	226
7.5	Conclusions.....	229
<b>CHAPTER 8</b>	<b>VOWEL DEVOICING AS A WEAKENING PROCESS IN JAPANESE</b>	<b>231</b>
8.1	Introduction.....	231
8.2	Experimental hypotheses.....	236



8.3	Experimental methods.....	237
8.3.1	Subjects.....	237
8.3.2	Experimental setting.....	237
8.3.3	Materials.....	238
8.3.3.1	Test words.....	238
8.3.3.2	Phonetic environments for the test words.....	239
8.3.3.3	Speech rate setting.....	239
8.3.4	Intensity measurements.....	241
8.4	Results and discussion.....	244
8.4.1	Voicing rates.....	244
8.4.2	Differences in vowel intensity.....	245
8.4.3	The effect of speaking rate on vowel intensity.....	252
8.4.4	Intensity ratios of voiced vowels of the same kind in devoicing and non-devoicing environments.....	256
8.4.5	Quality of voiced high vowels in devoicing environments.....	261
8.5	Vowel devoicing processes.....	279
8.6	Conclusions.....	293
<b>CHAPTER 9 CONCLUSIONS</b>		<b>295</b>
<b>REFERENCES.....</b>		<b>300</b>
<b>APPENDIX</b>		
I	(Chapter 3).....	311
II	(Chapter 6).....	316

## ABSTRACT

The processes of vowel devoicing in Standard Japanese were examined with respect to the phonetic and phonological environments and the syllable structure of Japanese, in comparison with vowel reduction processes in other languages, in most of which vowel reduction occurs optionally in fast or casual speech. This thesis examined whether Japanese vowel devoicing was a phonetic phenomenon caused by glottal assimilation between a high vowel and its adjacent voiceless consonants, or it was a more phonologically controlled compulsory process.

Experimental results showed that Japanese high vowel devoicing must be analysed separately in two devoicing conditions, namely single and consecutive devoicing environments. Devoicing was almost compulsory regardless of the presence of proposed blocking factors such as type of preceding consonant, accentuation, position in an utterance, as long as there was no devoiceable vowel in adjacent morae (single devoicing condition). However, under consecutive devoicing conditions, blocking factors became effective and prevented some devoiceable vowels from becoming voiceless.

The effect of speaking rate was also generally minimal in the single devoicing condition, but in the consecutive devoicing condition, the vowels were devoiced more at faster tempi than slower tempi, which created many examples of consecutively devoiced vowels over two morae.

Durational observations found that vowel devoicing involves not only phonatory change, but also slight durational reduction. However, the shorter duration of devoiced syllables were adjusted at the word level, so that the whole duration of a word with devoiced vowels remained similar to the word without devoiced vowels, regardless of the number of devoiced vowels in the word.

It must be noted that there was no clear-cut distinction between voiced and devoiced vowels, and the phonetic realisation of a devoiced vowel could vary from fully voiced to completely voiceless. A high vowel may be voiced in a typical devoicing environment, but its intensity is significantly weaker than those of vowels in a non-devoicing environment, at all speaking tempi. The mean differences of vowel intensities between these environments were generally higher at faster tempi.

The results implied that even when the vowel was voiced, its production process moved in favour of devoicing. However, in consecutive devoicing conditions, this process did not always apply. When some of the devoiceable vowels were devoiced in the consecutive devoicing environment, the intensities of devoiceable vowels were not significantly lower than those of other vowels.

The results of intensity measurements of voiced vowels in the devoicing and non-devoicing environments suggested that Japanese vowel devoicing was part of the overall process of complex vowel weakening, and that a completely devoiced vowel was the final state of the weakening process. Japanese vowel devoicing is primarily a process of glottal assimilation, but the results in the consecutive devoicing condition showed that this process was constrained by Japanese syllable structure.

# Chapter 1

## Introduction

The aims of this thesis are to examine the processes of vowel devoicing and the nature of devoiced vowels in Standard Japanese and to investigate the mechanisms which control the vowel devoicing processes. Vowel devoicing is fundamentally a physiological phenomenon, i.e. it is a change of phonation from voiced to voiceless. However, its processes are very complex, involving various linguistic factors including accentuation, position in an utterance, syllable structure, speech style, speaking tempo, syntactic significance of a word, and dialect. Vowel devoicing is a classical issue and has been a popular topic for phonetic and phonological research in Japanese. However, to date analysis of vowel devoicing has generally been limited to either phonetic or phonological approaches, and many aspects of devoiced vowels and vowel devoicing processes have not been widely studied.

This thesis will be a descriptive study of vowel devoicing in Japanese. I will first study vowel devoicing in relation to some phonetic and phonological factors, and then investigate the significance of the mora and syllable in Japanese as it relates to vowel devoicing. By investigating vowel devoicing processes, the aim is also to better understand the sound system and features of speech rhythm and timing in Japanese.

Close examination of the experimental results found that in vowel devoicing there is not a clear cut distinction between voiced and voiceless with the realisation of a vowel in the devoicing environment varying from fully voiced to completely voiceless. It is also the case that not all high vowels in a typical devoicing environment became



voiceless, but some of them remained voiced. Neither phonetic nor phonological approaches to devoicing can satisfactorily explain these features of devoicing. Phonological theories in general treat the devoicing process as categorical, and therefore phonological approaches cannot account for optionality of devoicing. Phonetic approaches may generally be a better means of explaining the devoicing process, but they still fail to explain devoicing patterns in consecutive devoicing environments.

In order to fully explain vowel devoicing processes, a proposal will be advanced for a new approach to vowel devoicing, namely that vowel devoicing is part of a general process of vowel weakening, which does not involve centralisation of quality, but does involve durational and intensity reduction. The vowel weakening process occurs when the laryngeal gestures overlap and do not achieve the target movements, resulting in weakening of vowel intensity. However, this process is conditioned by the syllable structure: i.e. this weakening process occurs only when the preceding consonant of the weakened vowel can be resyllabified to an adjacent syllable. In other words, both phonetic and phonological factors need to be considered in order to understand the mechanisms of Japanese vowel devoicing.

## 1.2 Overview

Vowel devoicing is the phenomenon where a vowel, which is usually a voiced sound, becomes voiceless, so that there is no voicing at the point where there is supposed to be a vowel. In Standard Japanese, the most likely candidates for vowel devoicing are the high vowels /i/ and /u/ when they are surrounded by voiceless sounds: either between voiceless consonants, or preceded by a voiceless consonant and followed by a pause.

Vowel devoicing occurs in many languages other than Japanese, but it is usually treated as an optional fast speech or casual speech process and is associated with the vowel reduction process in most of these languages (see Chapter 2). The typical



candidates for vowel devoicing and reduction in most languages are unstressed high vowels or schwa, both of which have an inherently short duration, and are often adjacent to a voiceless consonant. The economising of speech gestures can result in the devoicing of vowels. It is natural for the vocal folds which are pulled apart for a voiceless consonant to remain apart during the production of a short vowel, resulting in a devoiced vowel (see Chapter 6, Section 6.1).

Despite the physiological features surrounding the devoicing of vowels, vowel devoicing has been mainly treated as a phonological process in Japanese. It is partly because Japanese vowel devoicing is not only present in fast speech or casual speech, but also occurs regularly in formal speech at a normal tempo, and devoicing is specified at a lexical level in the dictionary. Japanese vowel devoicing is also dialect sensitive. It is a typical characteristic in Eastern dialects of Japanese including Standard Japanese. Under identical phonological environments, devoicing occurs in some dialects but not in others.

The devoicing process was described as (1) by generativists (McCawley, 1975, Hasegawa, 1979, Vance, 1987, etc.). A phonological rule such as (1) specifies the phonological environment where the process occurs: i.e. it specifies the devoicing environment (which is a high vowel surrounded by voiceless consonants) and the resultant outcome.

(1)

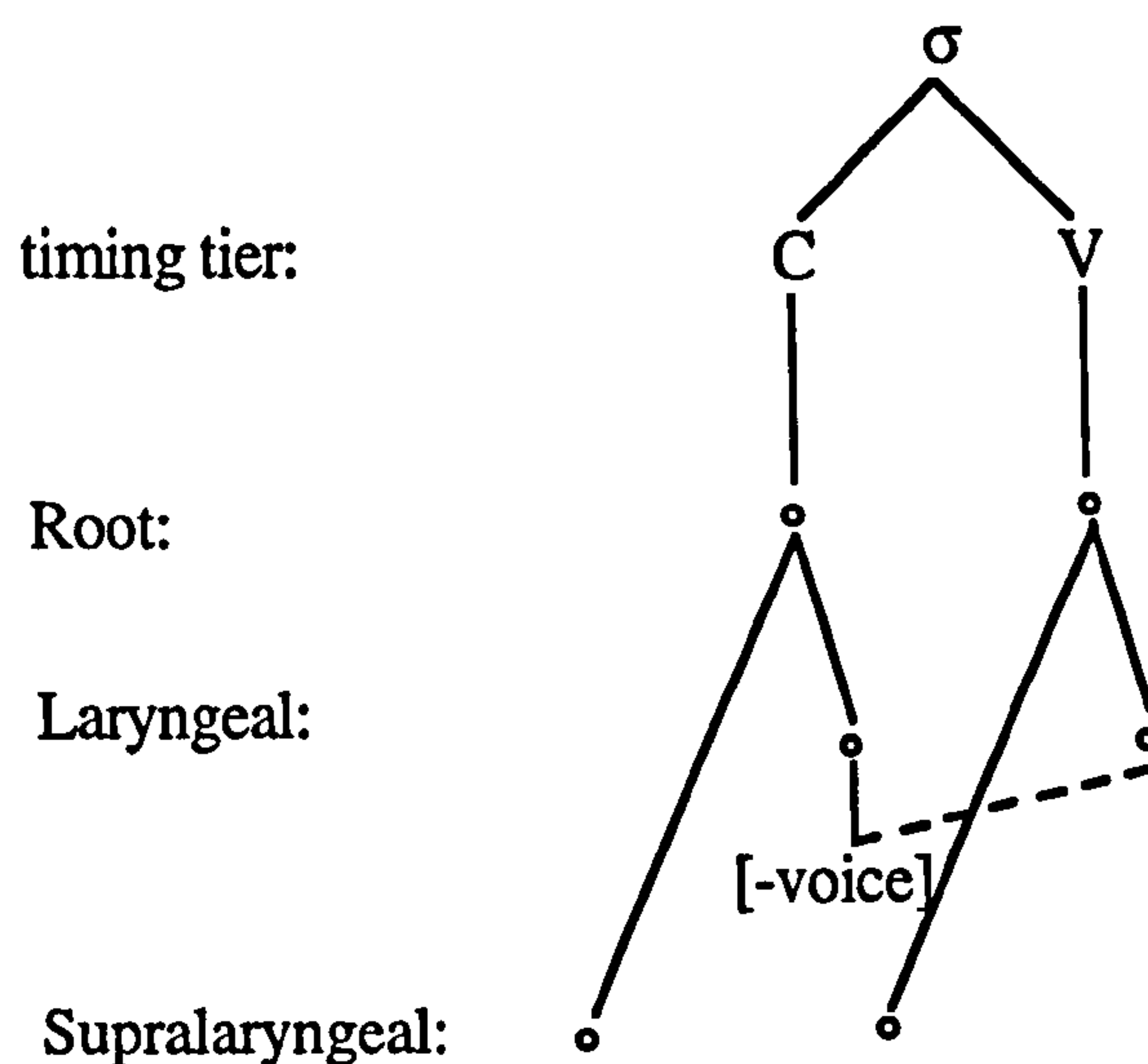
$$\begin{matrix} V \\ [+high] \end{matrix} \longrightarrow [-voice] \quad / \quad [-voice] \quad \longrightarrow \quad \left\{ \begin{matrix} [-voice] \\ \# \end{matrix} \right\}$$

Such categorical phonological theories, however, can only show the outcome of the vowel devoicing process without providing an explanation as to what causes the vowel to be devoiced in the first place.

Recently a new phonological approach based on autosegmental phonology has been proposed to explain vowel devoicing (Beckman, 1992). The autosegmental phonological representation of vowel devoicing as (2) can provide a phonetic explanation of how vowel devoicing occurs.

(2) (from Beckman, 1992)

" $\sigma$ " = syllable  
 "o" = timing node



In this theory it is proposed that devoicing is caused by laryngeal assimilation spreading the laryngeal feature of the preceding voiceless consonant to the following vowel.

However, this theory, as with (1) above is still a categorically based process. Both theoretical approaches are unable to explain fully the gradient nature of devoicing: e.g. the differences in the degree to which a vowel is devoiced. To the best of my knowledge, the degree of vowel devoicing has never been quantified. However, displays such as spectrograms, speech pressure waveforms and glottal waveforms show that the realisation of a devoiceable vowel is not a categorical distinction between fully voiced and completely voiceless, but that there are a lot of marginal cases, i.e. cases where vowels are partially (de)voiced. Vowels in the devoicing environment can be realised in any state from fully voiced to completely voiceless (see Chapter 5, Section 5.4.1.4).

In addition, devoicing is not absolutely consistent in any given environment. A word with a voiced vowel and its voiceless counterpart do not differ in meaning. For example, [aɕika] (with a voiced [i]) and [aɕika] (with a voiceless [j]) both mean 'sea lion', similarly [ka'sɯka] (with a voiced [ɯ]) and [ka'sɯka] (with a voiceless [ɯ]) both mean 'faint' (The symbol " ` " indicates pitch fall, or that the vowel immediately preceding the symbol is accented). Voiceless vowels are considered as allophones of their voiced counterparts.

Phonological rules also cannot account for the differences in devoicing rates depending on the type of preceding consonants. Some studies have reported an apparent tendency for preceding fricatives to trigger devoicing of the following high vowel more than preceding plosives (Han, 1962a, Takeda and Kuwabara, 1987, Vance, 1987, Yoshida and Sagisaka, 1990, etc.). However, phonological rules simply specify the devoicing environment, which is a high vowel surrounded by voiceless consonants, and its resultant outcome, but do not indicate the difference in devoicing rates between preceding fricatives and plosives.

It is also important to note that although Japanese vowel devoicing occurs in normal and slow speech, devoicing rates definitely increase in fast speech. If devoicing is purely a phonological process, an increase in speech rate should not affect the devoicing rates.

Recently, an alternative analysis suggested that vowel devoicing is a phonetic process, e.g. Jun (1993a) and Jun and Beckman (1993 and 1994). Their approach is based on a gestural score analysis by Browman and Goldstein (1989, 1990a, and 1990b) which uses phonetic evidence to build a phonological model (see Chapter 2, Section 2.4.4). A gestural overlap analysis not only explains the process and its outcome, but also the phonetic reasons behind the process. In this theory, vowel devoicing is fundamentally a laryngeal coarticulation between a high vowel and its adjacent voiceless sounds, both of which provide, phonetically, an ideal environment for the high vowel to become voiceless. The abducting gesture of the laryngeal muscles for the voiceless consonants



overlaps the adducting gesture of those for the short high vowel between the voiceless consonants. Because a high vowel is intrinsically short in duration when it is surrounded by voiceless consonants, it seems natural for the two laryngeal gestures to merge and result in a voiceless vowel. This theory not only provides a good explanation for the phenomenon but also explains the gradient nature of vowel devoicing. When the degree of overlap is great, the result will be complete devoicing, and when the overlap is slight, the result will be full voicing of the vowel. By gestural score analysis, Japanese vowel devoicing is a physiologically orientated process which occurs so regularly that it looks as if it were a phonological process.

The phonetic approach also provides a good explanation for the fact that, everything being equal, vowels are devoiced more often when they are preceded by fricatives than by plosives. Many physiological studies on the laryngeal muscles found that preceding fricatives create better laryngeal conditions for the vowel to become voiceless than preceding plosives (see Chapter 7, and Kuehn and Moll, 1976, Weitzman et al, 1976, Yoshioka et al, 1982, Simada et al, 1992, etc.).

However, there are some cases where gestural overlap analysis cannot provide an adequate answer, namely vowel devoicing in the consecutive devoicing environment: i.e. more than one high vowel surrounded by voiceless sounds in adjacent morae. In other words, under the same phonetic conditions, not all vowels in the consecutive devoicing environment undergo the devoicing process; some vowels are devoiced, but others are not. In Japanese, there are many phonological environments which offer possibilities for consecutive vowel devoicing. For example, all underlined high vowels in *kashitsuchishi* /kasitutisi/ ‘accidental death’ and *fukushika* /hukusika/ ‘social welfare department’ are devoiceable. Although consecutive devoicing has not been widely investigated, it is widely acknowledged that not all the devoiceable vowels undergo a devoicing process in normal speech: some of the devoiceable vowels become voiceless, but some remain voiced (Han, 1962a, Kawakami, 1977, Sakurai, 1985, Vance, 1987, Maekawa, 1989, etc.). If devoicing is a purely phonetic process, it is more natural for all these vowels to be devoiced. If there are no good phonetic explanations why some vowels remain voiced in a typical devoicing environment, it is

possible that there are other factors blocking the devoicing process in some of the vowels in the consecutive devoicing environment.

Previous studies have suggested that most factors which block vowel devoicing are phonological. These include accent on the voiceable vowel, presence of a following internal word boundary, and position in an utterance. On the other hand, most factors which promote vowel devoicing are phonetic. They include the type of vowels, the type of adjacent sounds especially the type of preceding consonants and speech rate. These facts suggest that although devoicing may primarily be a phonetic process, control of the overall mechanism of vowel devoicing is influenced by a combination of phonetic and phonological factors. Some processes are phonetically promoted, whereas others are phonologically constrained.

However, even the combination of phonological and phonetic analyses may still not be enough to explain the optionality of devoicing. This could possibly be explained if devoicing was interpreted as the ultimate stage of the vowel weakening process. Most previous studies on vowel devoicing used the binary distinction of voicing of vowels; namely voiced or devoiced. As stated earlier, a vowel in the devoicing environment does not always become completely voiceless, but is often shorter in duration and lower in intensity. If the degree of devoicing is also affected by the same controlling factors for overall devoicing, this suggests that vowel devoicing is actually part of a vowel weakening process and the ultimate stage of the weakening process is complete devoicing. In some languages such as Germanic languages, the vowel reduction process occurs in an unaccented position: vowels are centralised in quality and reduced in duration. The notable centralisation of vowels is said not to occur in Japanese, even in an unaccented position, though some variation of vowel quality occurs in reading style and spontaneous speech (Keating and Huffman, 1984). The Japanese vowel weakening process may operate differently from that in other languages. It does not involve reduction of vowel quality to the centralised position but involves the reduction of duration and intensity. The vowel weakening process occurs where the syllable structure permits it, and the realisation of the vowel can be at any stage from a fully voiced to completely voiceless vowel, though in most cases the vowel is realised as

completely voiceless. The proposal of vowel devoicing as a part of the process of vowel weakening could explain partial devoicing and the optionality of devoicing in the devoicing environment.

### 1.3 Outline of the thesis

This thesis consists of two main parts. Chapters 2 to 4 are literature reviews and Chapters 5 to 8 are experimental sections. Within the experimental section, Chapters 5 and 6 investigate vowel devoicing processes in relation to various factors of promoting and blocking devoicing. Chapters 7 and 8 examine the theory that vowel devoicing is part of the vowel weakening process, and attempt to understand the overall mechanisms of vowel devoicing in Japanese. Chapter 9 is a concluding chapter.

Chapter 2 is a review of the literature covering vowel devoicing in other languages. This chapter studies the processes of so called vowel devoicing and similar phenomena, i.e. vowel reduction and deletion in other languages, and defines these terms. Taking examples from various languages, common processes of vowel devoicing and common factors affecting the processes are discussed.

Chapter 3 is a brief review of Japanese phonology. This chapter highlights some areas of Japanese phonology which are relevant to vowel devoicing, and explains the relevance of the different devoicing factors which are investigated later in the thesis; namely mora and syllable, accent and intonation, vowels and consonants.

Chapter 4 is the review of past works on Japanese vowel devoicing. This chapter illustrates various factors affecting Japanese vowel devoicing in detail, presents and discusses areas of agreement and disagreement and the problems of research methods used in past studies. Based on the literature review in this chapter, the experiments for this thesis are planned. The main factor examined throughout the experimental chapters is the devoicing condition, i.e. the single and consecutive devoicing conditions. Since the consecutive devoicing environment is the only environment in



which devoicing does not always occur, studying the devoicing process in these two separate conditions may clarify what controls devoicing mechanisms.

Chapter 5 examines the effect of phonetic and phonological factors on vowel devoicing and the interaction between these two types of factors. The results of previous studies suggested that there are certain factors which promote vowel devoicing because they create phonetically ideal conditions for devoicing. On the other hand, there are some other factors which block vowel devoicing. In this chapter, I examine the effect of four factors, namely (a) type of preceding consonants, (b) accentedness, (c) presence of a following internal word boundary, and (d) position in an utterance. The aims of the experiment are to investigate (i) how devoicing in a phonetically ideal environment (a vowel preceded by a fricative) is affected by the blocking factors for devoicing (the presence of an accent and a following internal word boundary, and the position in an utterance), and (ii) the effectiveness of the blocking factors in the two devoicing conditions, namely single and consecutive devoicing environments.

Chapter 6 investigates the effect of speaking rate on vowel devoicing and also the interaction between factors promoting and blocking devoicing. This experiment examines the main factor promoting devoicing namely an increase in speaking rate, in relation to type of preceding consonant, position in an utterance in two devoicing conditions, i.e. single and consecutive devoicing environments. Vowel devoicing in most languages seems to occur more at a fast tempo (see Chapter 2). The questions to be asked are whether (i) devoicing occurs more often at fast tempo as in other languages, (ii) slow tempo blocks vowel devoicing, and (iii) if there is any tempo effect, is it effective in both devoicing conditions.

Chapter 7 studies the temporal aspect of devoiced vowels. The experiment examines the duration of a mora when a vowel is devoiced, in other words, whether durational reduction is related to the Japanese vowel devoicing process. Vowel devoicing processes in other languages involve vowel reduction in duration, intensity and quality. Based on the results presented, the status of mora and whether the phenomenon is really devoicing or deletion is also discussed.

Chapter 8 investigates the reduction of intensity and quality of devoiced vowels. The experiment examines whether voiced devoiced vowels are phonetically the same as vowels in the non-devoicing environment. The degree of voicing will be quantified by measuring the intensity of vowels in the voicing and devoicing environments and also in the single and consecutive conditions. The results are discussed in relation to the overall devoicing process, and whether Japanese vowel devoicing is similar to the vowel reduction process in other languages. Finally, I will propose controlling factors for the overall mechanisms of vowel devoicing, and discuss the significance of the mora and syllable in Japanese. The mora has been considered as the primary unit of Japanese, and the status of the syllable has been regarded as less important (see Chapter 3). I will attempt to establish the importance of syllable as well as the mora in Japanese.



# Chapter 2

## Cross-linguistic study of the process of vowel devoicing

### 2.1 Introduction

The phenomenon called vowel devoicing<sup>1</sup> has been reported in many languages. Before examining vowel devoicing in Japanese, in this chapter I will discuss similar phenomena in other languages. There are quite a few common features in vowel devoicing across languages, and there are also some language specific features. Phonetic, phonological and sociolinguistic factors all affect vowel devoicing. The term "vowel devoicing", in practice, includes "vowel reduction" and "vowel deletion". Both vowel devoicing and vowel reduction involve a phonetic change of vowels, but vowel deletion can involve phonological change. Both intrinsic features of vowels such as inherent duration of each vowel and extrinsic factors are involved in vowel devoicing. Extrinsic factors include phonological environments, stress and accent, speaking tempo, lexical, syntactic and semantic factors, socio-linguistic factors such as speaking style, generation and sex difference.

The vowel devoicing process varies from language to language. In some languages, the process is physiologically determined, and in others the process reflects other phonetic and phonological processes. In some cases, the devoicing process is

---

<sup>1</sup> Some linguists use the term "unvoicing" instead of "devoicing", especially earlier works. However, the terms "unvoicing" and "devoicing" seem to refer to the same phenomenon. Therefore, I use the term "devoicing" in this thesis.

constrained by other phonological factors such as the syllable structure (see Sections 2.2.2).

In this chapter, I will first define "vowel devoicing", then discuss both language specific and universal vowel devoicing factors and devoicing processes.

## **2.2 The definition of vowel devoicing**

### **2.2.1 Vowel devoicing, reduction, deletion and weakening**

As stated earlier, the definition of the term 'vowel devoicing' is not very clear. It is often confused with other related phenomena, namely 'vowel reduction' and 'vowel deletion'. These terms are used interchangeably in some studies when discussing the same phenomenon, although strictly speaking, they are different linguistic phenomena. Therefore, before discussing vowel devoicing in various languages, I will initially define these terms.

Vowel devoicing is a phonatory change of a sound from voiced to voiceless. When a vowel is devoiced, it is still possible to trace the presence of the vowel acoustically, or the vowel is still identifiable auditorily in both duration and perception, because a devoiced vowel has the resonance of its voiced counterpart. More importantly, a devoiced vowel does not lose its status in a syllable. 'Vowel reduction' usually means centralisation of quality, but often involves durational reduction and intensity reduction. 'Vowel deletion' is a different process from devoicing and reduction. I define vowel deletion as where a vowel disappears auditorily and acoustically leaving no trace of its presence. As a result, vowel deletion causes resyllabification, since a syllable with a deleted vowel loses its syllabicity. Deletion by strict definition is much rarer than devoicing or reduction. A vowel may be auditorily deleted but actually be present acoustically retaining its duration, or a vowel may be durationally deleted but retain its presence as a coarticulatory effect on its neighbouring sounds. These cases are not examples of vowel deletion.

All of these three phenomena are 'vowel weakening' processes. I define the term 'weakening' in this thesis as a superordinate term covering any of the aspects of the above phenomena, rather than the general usage of the term meaning vowel lenition, because all the processes involve some aspect of weakening. Vowel devoicing involves weakening of sonority. Vowel reduction involves not only vowel centralisation, but often also weakening of the intensity and duration of the vowel. Vowel deletion involves the loss of sonority and syllabicity. All of these make up the overall process of weakening.

However, there are not always clear cut distinctions between the three definitions. These are not totally independent processes, as will be shown in examples later in this chapter. There are significant areas of overlap between these processes. For example, as described below, vowel devoicing can occur independently, but can also be a consequence of the vowel reduction process. It is often the case that a vowel is first reduced, and then eventually becomes voiceless. Similarly, vowel deletion may occur independently, but often it occurs following other processes: a vowel is reduced first, then devoiced and finally may be deleted. Therefore, vowel devoicing has to be considered in relation to vowel reduction and deletion.

Moreover, a process which may be regarded as 'deletion' is often treated as 'devoicing' in order to explain phonological processes. Vowel deletion causes various phonological changes such as resyllabification and accent shift. Therefore, treating the process as 'devoicing' may be advantageous, since a 'devoiced' vowel can retain its phonological status such as syllabicity and accentuation. In the next section, I will present various phonological processes which are normally described as vowel 'devoicing'.

### **2.2.2 Phonological processes of vowel devoicing**

Although the strict definition of vowel devoicing is the change of phonation from voiced to voiceless, vowel devoicing is in practice a more complicated phenomenon



and has to be considered with other factors such as syllabicity, accentuation, phonotactics and coarticulation. For example, in cases where auditorily and/or acoustically there is no vowel, but the syllable retains its syllabicity and/or accent, this may be treated as devoicing rather than deletion. Similarly a vowel is not considered deleted, but rather devoiced, (1) if deleting the vowel violates the phonotactics of the language in question, or (2) if there is a coarticulatory effect of the 'deleted' vowel on its adjacent sounds.

The phonetic features of a vowel can affect surrounding consonants. In natural speech, coarticulation between neighbouring segments is inevitable. For instance, a high front vowel has a palatalising effect on neighbouring consonants. Even when a high front vowel is phonetically not present, it may still be possible to predict the feature of the underlying vowel manifested in the palatalisation of a preceding consonant. Similarly, although a high back vowel is not present in terms of oral articulation, it may still give a velar feature to a preceding consonant. If phonetic features of remaining segments allow the prediction of an underlying vowel, it is not strictly appropriate to say that a vowel is 'deleted'.

Vowel devoicing and reduction do not normally alter syllable structure, but reduction may involve reorganisation of syllable structure. One of the examples is syllabic consonants. Syllabic consonant formation is a process of syllable reduction. The oral articulation of a vowel is not present in a syllable which consists of only a consonant, but the syllabicity of the vowel is acoustically manifested as the syllabicity of the consonant. The duration of syllabic consonants is longer than that of non-syllabic consonants. However, similar to syllables with reduced vowels, syllabic consonants are inherently shorter and do not have the same status as full syllables. In English, for example, a syllable with a schwa may be reduced to a syllabic consonant: e.g. *sparkle* ['spɑ:kəl] ---> ['spɑ:kɫ], and *battle* ['bætəl] ---> ['bætɫ]<sup>2</sup>.

---

<sup>2</sup> The schwa may be deleted completely and the syllabic consonants may be desyllabified when a weak vowel follows them as shown in (1).

(1)	<i>sparkle</i>	['spɑ:kɫ]	--->	<i>sparkler</i>	['spɑ:klə]
	<i>battle</i>	['bætɫ]	--->	<i>battling</i>	['bætɫɪŋ]

Extremely reduced vowels are often treated not as 'deleted' vowels but as 'devoiced' vowels. Such cases are (a) when retaining a vowel as a devoiced vowel may maintain the phonological structure, such as the phonotactics and the syllable structure, of the language in question, or (b) when blocking a certain phonological process may be explained by the presence of the devoiced vowel.

For example, in Toposa, a Teso-Turkana language<sup>3</sup> spoken in Sudan, there are 9 vowels which are divided into two groups, namely (a) advanced tongue root /i, e, o, u/, and (b) non-advanced tongue root /ɪ, ɛ, a, ɔ, ʊ/. All the vowels have voiceless counterparts (Schröder and Schröder, 1987: 17). The presence or absence of a voiceless vowel changes the meaning of words. There are many examples of word final voiceless vowels with little or no phonetic realisation, and native speakers of Toposa seem to have difficulty in recognising voiceless vowels but realise that there is something present. The phonetic realisation of voiceless vowels varies between (i) no realisation, (ii) modification of the preceding consonant, (iii) whisper, and (iv) fully voiced, depending on the environment in which they occur. Words which end with a consonant are phonologically transcribed with voiceless vowels; in other words, no Toposa words end with a consonant, therefore words end either with a voiced vowel or a voiceless vowel. This is because (a) the basic syllable structure of Toposa is CV, and a vowel is inserted in cases of consonant clusters in loan words, (b) historically there was a vowel in word final position since other related languages retain a voiced vowel word finally. (Schröder and Schröder, 1973).

Griffen (1985) presented examples of coarticulatory effects and retained the tone and syllabicity of underlying vowels in Mandarin Chinese in which many instances of syllables consisting of obstruents without a vowel occur; such as [sʒ] 'four' with falling tone over a syllable that shifts directly from a voiceless fricative to a voiced fricative without any periodicity without oral constriction. It would not be too much of a problem if sonorants carried the prosodic features as they could become syllabic under some circumstances, e.g. between obstruents. However, obstruents which

---

<sup>3</sup> All Teso-Turkana languages have voiceless vowels (Schröder and Schröder, 1987: 17).

have very weak sonority do not normally occur at the peak of sonority of a syllable. In syllables consisting only of obstruents, it is possible to trace the underlying high back unrounded vowel /ʊ/, not only because the precise articulatory points such as movements of the tongue body, jaw, lips, etc. are constrained by the underlying vowel, but also because of the fact that the syllable prosody of the falling tone is realised throughout the syllable. The fricative coarticulates with the vowel; i.e. the consonant occurs simultaneously with the falling tone of the vowel which dominates the entire length of the syllable. Therefore, syllables with no vowels such as [sʒ] ‘four’ and [dʒ] ‘word’ may be analysed as /szʊ/ and /dzʊ/ respectively.

Another example can be seen in Swabian, a west Germanic dialect spoken mainly in the German state of Baden-Württemberg to the north of Switzerland. In Swabian, the presence or absence of a vowel determines the application of certain phonological rules, and therefore in certain environments vowels are described as voiceless although they are phonetically not present (Griffen, 1983). For example, Swabian has a wide spread rule of strengthening consonants in word initial position, i.e. from lenis to fortis, through vowel devoicing. Thus Middle High German (MHG) *behalten* ‘given’ is in Swabian realised as [pʰaltə] and *gehebt* ‘lifted’ (MHG) is [kʰet], and *gelogen* ‘lied’ (MHG) is [klɔgə]. However, words such as *glücke* ‘luck’ (MHG) remains [glik] and *geluste* ‘desire’ (MHG) is [glʊfdə] in Swabian, with no strengthening of the initial consonants. Griffen (1983) explained the presence of the devoiced vowel as the reason for the change from lenis to fortis consonants.

In Swabian, the prefixes *be-*, *ge-*, and *de-* normally involve extremely weak articulation [bə], [gə] and [də] respectively. The weakly articulated vowels are devoiced and shortened, then devoice a preceding consonant, which may become syllabic or coalesce with the aspiration of the following /h/ and become [pʰ], [tʰ] and [kʰ]:

- (2) a. [bə] ---> [bə] ---> [b̥] ---> [b] ---> [p] (---> [p] ---> [pʰ])  
 b. [də] ---> [də] ---> [d̥] ---> [d] ---> [t] (---> [t] ---> [tʰ])  
 c. [gə] ---> [gə] ---> [g̥] ---> [g] ---> [k] (---> [k] ---> [kʰ])



/be-/ and /ge-/ in *behalten* 'given', *gehebt* 'lifted' and *gelogen* 'lied' are prefixes and thus /ə/ in these words undergoes the devoicing process, and eventually the initial voiced lenis consonants are strengthened to voiceless fortis counterparts. In *glücke* 'luck', however, there is no /ə/ which undergoes the weakening process and /ge-/ in *geluste* 'desire' is not participial but lexical, thus it is not weakly articulated. Although the phonetic realisations in Swabian [p<sup>h</sup>altə], [k<sup>h</sup>ɛt], and [kløgə] do not have schwa, it exists phonologically as a voiceless vowel which causes the change of the initial voiced consonants to voiceless fortis.

Vowel deletion in a strict sense is not very common, and deletion is often confused with devoicing or reduction. As vowel deletion involves a loss of syllabicity and vowel devoicing involves loss of voicing and intensity reduction, both vowel deletion and devoicing can be considered as part of the overall vowel weakening process. In many languages, in fact, as illustrated above 'deleted' vowels are often treated as 'devoiced' vowels in order to preserve the phonotactics and syllable structure of the languages. In this chapter, I will present examples of all aspects of the vowel weakening process, namely vowel reduction, devoicing and deletion.

### 2.2.3 Phonetic description of voiceless vowels

In addition to the vowel devoicing processes described above, devoicing of a vowel occurs as inevitable phonetic consequence of various processes. Vowels are phonetically defined as sounds which are articulated without a complete closure in the mouth, or a degree of narrowing which would produce audible friction; the air escapes evenly over the centre of the tongue (Crystal, 1985: 330). The term 'vowel devoicing' implies that all vowels are voiced at some level. It is taken for granted that vowels are voiced. Therefore, the description of a vowel does not normally specify the state of the glottis, and a vowel usually has a three-dimensional description, i.e. [±high], [±back] and [±round]. In addition to that, words meaning 'vowel' in European languages such as the English word 'vowel', French 'voyelle' and Spanish 'vocal'

came from the Latin word 'vocalis' which means 'vocal' , a derivative of 'vox' ([the] voice) (Abercrombie, 1967: 59, Donegan, 1985: 46).

The majority of vowels spoken in world languages are in fact voiced. However, vowels can be produced with types of phonation other than voiced. Catford (1977: 96) stated that "in order to be audible at all, a voiceless vowel must have turbulent flow through one of its orifices: the oral orifice for close (approximant) vowels like [i, u], the glottal orifice for open (resonant) vowels like [ɛ, a]".

In terms of phonatory stricture-types, Catford distinguishes eight types, namely voiceless, whisper, voice, creak, and combinations of these phonatory types, i.e. breathy voice, whispery voice, whispery creak and creaky voice (Catford, 1977: 95-101). During the production of a voiceless sound, the glottis is wide open and there is either little or no airflow (nil-phonation), or there is a considerable amount of airflow generating the sound of audible breathing (breath phonation). For whisper, the glottis is open but considerably narrowed, generating a strong, rich, hushing, turbulent sound. During the production of a voiced sound, the vocal folds are close together and are subjected to varying degrees of tension, and they vibrate, generating the periodic sound. In whispered speech, normally voiced sounds are whispered, while normally voiceless sounds remain voiceless.

Phonologically, vowel devoicing is a categorical process; either a vowel is voiced or devoiced. But phonetically it is a gradual process extending from fully voiced to complete deletion, which may be closely dependent on speaking style and rate. In theory, there is a continuous change of phonatory state from full voicing to complete devoicing. Like voiced consonants, (voiced) vowels can be completely or partially devoiced. For a fully voiced vowel, the vocal folds keep vibrating from the onset to the offset of a vowel. For complete vowel devoicing, the state of the vocal folds for voiceless sounds lasts throughout the production of the vowel. For partial devoicing, either the onset of vocal fold vibration is delayed or the vocal folds stop vibrating in the final part of a vowel. An intermediate phase between fully voiced and completely voiceless is whisper. Catford (1977: 96-97) defined whisper phonation as where the



critical velocity of airflow is very much lower than voiceless phonation, since the glottal orifice for whisper is very much smaller than that for voiceless phonation. Catford also stated that theoretically various degrees of glottal opening, from breathing position (the widest) to strong whisper (the narrowest), are possible, but in practice there are only two possibilities: (1) the wide open glottis of 60-95 % of the maximal glottal area for voiceless, and a very narrow one of less than 25 % of the maximal glottal area for whisper.

A vowel with either voiceless or whisper phonation is considered to be a 'devoiced' vowel in the description of languages. However, 'devoiced' should be differentiated from 'voicelessness'. As far as consonants are concerned, 'devoiced' voiced consonants remain lenis whereas voiceless consonants are said to be fortis (Catford, 1977: 112, Gimson, 1980: 287). Fortis consonants are said to be tense, whereas lenis consonants are lax. In general, voiceless consonants are said to be fortis and voiced consonants are said to be lenis. For instance, Kim (1965: 340) cited French examples from Jakobson and Halle (1963) and demonstrated that *vous la jetez* /vu la ʒte/ [vu la ʒte] 'you throw it' and *vous l'achetez* /vu laʃte/ [vu laʃte] 'you buy it' are a minimal pair, in which both [ʒ] and [ʃ] are voiceless, but [ʒ] is said to be lax and thus still distinguished from [ʃ] which is said to be tense. Another example is found in English. In English, so called 'voiced' obstruents are partially or fully devoiced when they are adjacent to a voiceless sound or a pause. However, 'devoiced' voiced obstruents remain lenis whereas voiceless obstruents are fortis (Catford, 1977: 112, Gimson, 1980: 287).

Vowels can be partially devoiced in certain environments in some languages. For example, Swedish voiced obstruents have similar characteristics to the voiced obstruents in many other Germanic languages: they are partially or fully devoiced in utterance initial and final position. The experiments of Gobl and Ní Chasaide (1988) showed that in Swedish, when a voiceless consonant followed a vowel, the phonatory mode of the first vowel in /bap:a/ became breathy voiced as it approached the voiceless geminate /p:/, as opposed to the vowel /a/ before the voiced geminate /b:/ in /bab:a/ and /vab:a/. For some speakers the voicing of the vowel ceased completely after about 50

ms out of the total vowel duration of about 200 ms prior to the oral closure for /p/ and there was a short interval of true voicelessness.

The initial devoicing of a vowel could also be interpreted as a delay of voice onset time after voiceless consonants, especially stops. In languages such as Thai and Germanic languages except for Dutch, voiceless stops are aspirated<sup>4</sup> in syllable initial position, and the aspiration is manifested as the delay of voice onset of the following vowel (Catford, 1977: 112, Gimson, 1980: 153, Ladefoged, 1982: 132).

Normally vowels are produced with a pulmonic airstream mechanism, but can be made with a glottalic or velaric airstream mechanism. In these cases, if the articulators are in the position for a vowel when a glottalic or velaric mechanism is brought into play, the result is inaudible (Abercrombie, 1967: 60). When [h] is followed by a vowel, it is in fact a voiceless version of the following vowel (Abercrombie, 1967: 169).

Laver (1994: 342) stated that “when considering the onset and offset of segments internal to an utterance, one segment’s offset transition overlaps with the onset of the next segment. This could result in confusion in particular cases about whether devoicing of an individual segment was full or partial, unless some convention about phasing is agreed. The boundaries of the medial phase of approximants and other vocoids are less distinctly identifiable than those in stops and fricatives”. Fant (1960: 207) also stated that “segmental boundaries can be specified in terms of temporally localised changes in the two basic categories of sound production, the source and the filter. These changes are: (a) change of the type of source (voice/noise), (b) intensity of the source (amplitude curve), (c) rapid change in the vocal tract filter function (formant transition), (d) simultaneous changes in both filter and source”. Laver (1991: 209) defined the full voicing of a segment thus: “for a stop or a fricative to have full

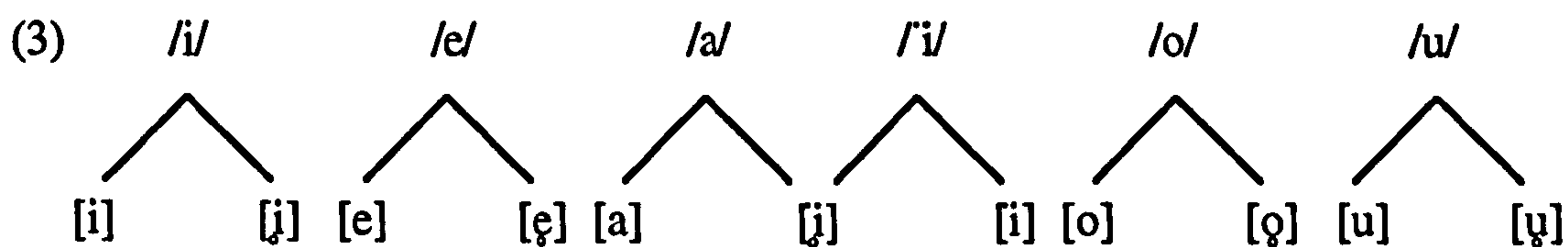
---

<sup>4</sup> Voiced aspirated stops occur in many languages spoken in India. They are followed by a vowel pronounced with breathy voice (Abercrombie, 1967: 149). During the sequence of ‘voiced aspirated stop + vowel’, vocal fold vibration does not cease throughout the sequence. Voiceless fricatives are also aspirated in Korean (C.-W. Kim, 1965 and 1970). In the case of pre-aspiration, which occurs in Scottish Gaelic, Icelandic, Faroese, Finnish and Lappish, a devoiced version of the preceding vowel precedes a pre-aspirated consonant (Thráinsson, 1978 and 1980, Árnason, 1986, Laver, 1994: 358).

voicing, voicing must be in evidence throughout the medial phase of the segment, i.e. any devoicing of stops or fricatives has to involve a loss of voicing during some part of the medial phase of the segments. On the other hand, for syllable-marginal approximants and all syllable-nuclear vocoids except for monophthongs, the active articulators are in continuous movement from onset to offset of the segment, i.e. there must be voicing throughout the onset, medial and offset of the segment". Therefore, a partial devoicing of vowels caused by aspiration of an adjacent consonant, and by voicing assimilation to the following consonant as in the above case in Swedish is said to be vowel devoicing.

#### 2.2.4 Are voiceless vowels distinctive?

It is generally believed that voiceless vowels are not distinctive. The environments where voiceless vowels occur can be specified phonologically in most languages, which suggests that they are allophones of voiced counterparts. Voiceless vowels and their voiced counterparts are not normally in complementary distribution, but voiceless vowels are considered to result from optional speech processes. Very few languages treat voiceless vowel as phonemes, but they may be contrastive to voiced counterparts in certain environments in some languages. For example, Comanche has six vowels and all of them have a voiceless allophone. As listed in (3), vowel features other than voicing are usually preserved in these voiced-voiceless allophones, except that the low vowel [a] most often alternates with the high front vowel [ɨ] rather than the expected [ʌ].



Each pair of voiced and voiceless vowels is contrastive in the examples in (4).





(5)	/nókohkiʔuʔ/	/háʔni/
	↓	↓
by rule (i)	*[nókq̥hkiʔuʔ]	-----
by rule (ii)	[nókq̥kiʔuʔ]	[ʔáʔni]

In summary, it seems that voiceless vowels are not distinctive in any language and are never considered as separate phonemes.

2.3 Factors affecting vowel reduction, devoicing and deletion

2.3.1 Intrinsic features of vowels

In most languages, there is a tendency for certain types of vowels to be devoiced: i.e.(1) a short or weak vowel, (2) a vowel which is identical to a nearby vowel, and (3) a vowel modified by surrounding consonants (Hooper, 1972: 141). Vowels which undergo devoicing processes are predominantly high vowels and schwa, although other vowels can be devoiced in some languages, e.g. Continental Portuguese, Toposa, Southern Paiute, Shoshoni, Comanche and Cheyenne<sup>6</sup> (Johns, 1972, Schröder and Schröder, 1987, Casagrande, 1954, Canonge, 1957, Ogawa, 1981, Armagost, 1986). This is probably due to inherent qualities of these vowels. Vowel devoicing seems to be partly dependent on the intrinsic duration of the vowels, i.e. shorter vowels tend to be devoiced in certain environments.

Lehiste (1970: 18) stated that as far as the vowels were concerned their duration appeared to be correlated with tongue height: other factors being equal a high vowel was shorter than a low vowel. Elert (1964) quoted by Lehiste (1970: 18) found that for short allophones in Swedish, if the average duration of high vowels /i, y, o/ is taken as 1.00, that of mid-vowels /e, u, ʌ/ was 1.08 and that of low vowels /ä, ö, a/ was 1.17. This tendency seems to be a language universal.

<sup>6</sup> In these languages, all vowels have a voiceless allophone.

Southern Paiute	5 vowel system	/i, i̥, a, o, u/
Shoshoni	6 vowel system	/i, i̥, e, a, u, o/
Comanche	6 vowel system	/i, i̥, e, a, u, o/
Cheyenne	3 vowel system	/e, a, o/

Keating (1985: 118) referring to Lindblom (1967) stated that the observed differences in vowel duration could be accounted for by automatic biomechanical effects rather than by deliberate temporal control. Referring to the result in Westbury and Keating (1980), Keating (1985: 119) also made a similar comment on intrinsic vowel duration that lower vowels were longer in acoustic duration than higher vowels. The vowels with the lower jaw position had longer acoustic duration. The longer duration was due to longer travel times to reach the target positions, not longer steady states of the low vowels. Keating concluded that physical factors definitely influence vowel duration but do not completely control it since different languages have different vowel durations. Gay et al. (1974), Gay (1978) and Kuehn and Moll (1976) demonstrated that longer duration is not the inevitable consequence of slow articulation; speakers can increase the velocity of articulatory movement depending on the distance (for further detail, see Chapter 6, Section 6.1).

Another vowel involved in devoicing is schwa [ə]. Similar to [i] and [u], schwa has the intrinsic feature of short duration, and it is not normally stressed. Schwa is very often a product of vowel reduction or centralisation. In some languages, vowels are reduced in an unstressed syllable. Vowel reduction is a vowel weakening process: one step towards deletion. Historically, in English and French, final unstressed vowels weakened to schwa and then were dropped (Hyman, 1975: 170, Gimson, 1980: 127). It is also possible to say that schwa is a product of destressing and neutralisation of other vowels. Therefore, schwa is inherently weak and unstressed<sup>7</sup>. Further details of vowel reduction and centralisation will be discussed in Sections 2.3.2 and 2.4.

In languages such as English and German, schwa often loses its vowel status and makes an adjacent consonant syllabic. For example:

---

<sup>7</sup> It is possible to have phonemic stressed schwa in some words, such as in RP *bird* /'bɜ:d/ or in some American dialects *bud* /'bəd/.





## 2.3.2 Extrinsic factors

### 2.3.2.1 Phonological environments

A typical environment for vowel devoicing is being adjacent to a voiceless consonant. In most languages, vowel devoicing occurs between voiceless consonants, but in some languages such as some Amerindian languages (Comanche, Shoshoni, Southern Paiute and Cheyenne), and Continental Portuguese, vowels are devoiced word- or utterance-finally before a pause (Casagrande, 1954, Canonge, 1957, Ogawa, 1981, Armagost, 1986).

In Greek, an unstressed high vowel is most likely to be reduced when surrounded by voiceless consonants /p, t, k, f, θ, s, x/, and subsequently is often devoiced. The second most likely environment for high vowel reduction is when one of the adjacent consonants is voiceless. In this case, the preceding voiceless consonant is more influential than the following voiceless consonant. Although it is possible, vowel reduction between voiced consonants is very rare (Dauer, 1980b).

Some languages require more complicated rules for the devoicing environments. For example, in Comanche, too, it is essentially adjacent voiceless consonants that trigger vowel devoicing, but the devoicing environment requires further conditions. It has been suggested that stridency of preceding and following consonants blocks devoicing. The vowels are devoiced when they occur before /h/ and /s/, but when a vowel is preceded by /c/, a following /h/ triggers devoicing but a following /s/ does not (Armagost, 1986).

Voiceless vowels can occur when adjacent to voiced consonants. In Toposa, voiceless vowels occur after voiced consonants /b, d, g, m, n, ɲ, r, l, j, w, ʝ/ before a pause. However, vowel devoicing is rare between voiced consonants.

The intrinsic durational feature of vowels mentioned in Section 2.3.1 is also influenced by other extrinsic factors. Chen (1970) demonstrated the more difference in vowel duration when vowels are followed by voiceless consonants than by voiced



consonants in English, German, French, Korean, Norwegian, Russian and Chinese. Lehiste (1970: 27) found a similar difference in Spanish.

Although the effect of following consonants on the duration of vowels seems to be language universal, the degree of the effect is language specific, depending on the phonological structure of the languages in question. The effect of following consonants on vowel reduction is greater in languages such as English, but smaller in languages such as French, German and Norwegian (Chen, 1970: 139). The results of Keating (1985: 121) suggested that Polish and Czech as well as other West Slavonic languages did not show a significant difference of vowel duration before the voiced-voiceless contrast. The duration of vowels is determined by some contextual differences such as adjacent consonants, but language specific factors such as stress and rhythm are very influential, and these language specific factors balance the intrinsic syllable duration in the language.

### 2.3.2.2 Stress and accent

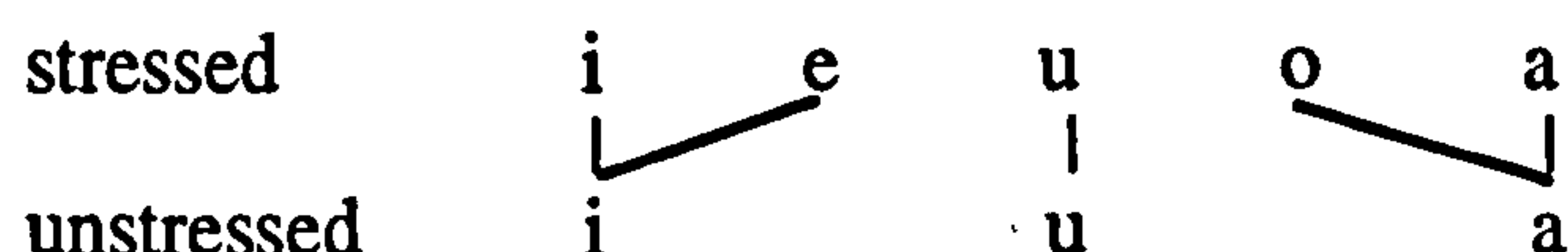
Different analyses are required for vowel reduction depending on the nature of the language. Vowel reduction is a typical characteristic of languages in which stress plays an important role, such as English, German and Swedish (Lindblom, 1963: 1773). Languages such as French, Japanese and Spanish, in which each word has a more or less fixed stress pattern, do not have obviously reduced allophones of vowels in unstressed position as opposed to in stressed position (Abercrombie, 1967: 98, Dauer, 1980a: 338, Keating, 1984: 196). In languages such as English, German and Swedish, most vowels in unstressed syllables tend to be neutralised to schwa in quality (Lindblom, 1963: 1773). Citing from Stetson (1951), Tiffany (1959), Shearme and Holmes (1962) and Fant (1962), Lindblom (1963: 1773-1774) concluded that in British English, American English and Swedish, as the stress decreases, which usually relates to the duration of the vowel, vowels shift to a neutralised schwa<sup>10</sup>.

---

<sup>10</sup> In English, however, unstressed /i, ɪ, u/ do not reduce very much (Gimson, 1980: 146-148).

In Russian, which is another language in which stress plays an important role, formant frequencies of unstressed vowels differ greatly from those of the corresponding stressed vowels (Lindblom, 1963: 1774 quoting Jones, 1959). Some Russian dialects have a neutralisation rule for certain vowels in unstressed syllables (Lehiste, 1970: 141). There is a 5-vowel contrast in stressed syllables, but a 3-vowel contrast in unstressed syllables:

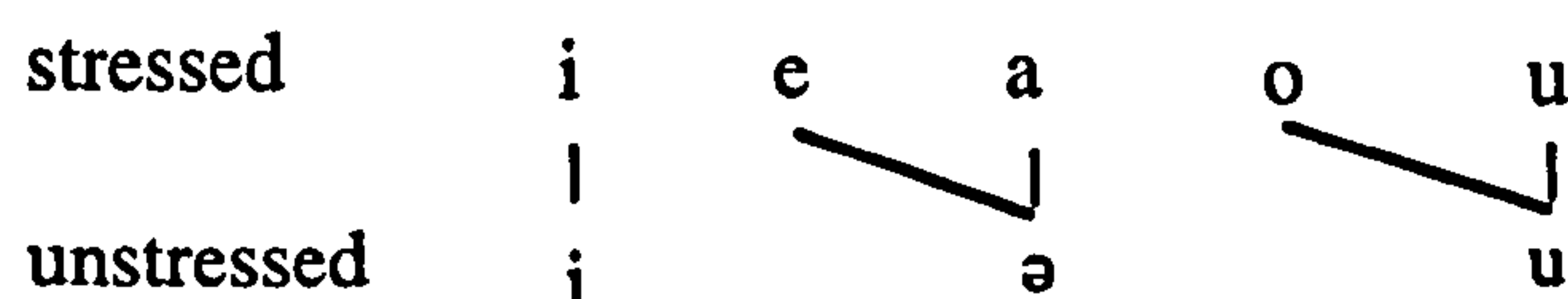
(8)



(Hyman, 1975: 169)

Catalan also has five vowels in stressed syllables but three in unstressed syllables.

(9)



(Wheeler, 1979: 3 & 55)

Similarly in modern Greek, stressed /á/, /é/, /ó/ have more open peripheral positions than unstressed /a/, /e/, /o/, and /í/ and /ú/ have closer centralised positions than /i/ and /u/ (Mirambel, 1948: 18, 1959: 27, quoted by Dauer, 1980a: 80).

Centralisation of vowels is contextual assimilation (Lindblom 1963: 1781). The effect of stress on vowel duration is language specific, dependent on its function, and importance and the structure of the language in question (Dauer, 1980a: 85).

Another typical environments for vowel weakening are the pre-tonic and post-tonic positions. In the process of vowel deletion in Continental Portuguese, in which three vowels /i, a, u/ may be deleted, stress plays an important role. Vowels in any unstressed word final syllable of CV structure followed by a syllable without a pause

intervening can be deleted. The typical environments are where: (a) word final syllable and a following word initial syllable are both weak, and (b) word final syllable is weak and a following word initial syllable starts with a vowel and is stressed (Johns, 1972). Either a word final unstressed vowel may coalesce with a word initial unstressed vowel of a following word, or a word final unstressed vowel may merge with a word initial stressed vowel of a following word. In both cases, there is a reorganisation of syllable structure. For instance;

- (10) a. /i:/ /'devi a'ver/                      ['deevə 'veɹ]                      'there must be'  
           /a:/ /'mɛrə 'bɪsili'dadi/                      ['mɛɹɪmbʃili 'ðɑɹɪ]                      'pure idiocy'  
           /u:/ /'ɪstu ɛ 'bɒ/                      ['ɪstɛ 'βɒ]                      'this is good'  
       b. /i:/ /'eu kuazi 'iij/                      ['eu kwə 'ziɹə]                      'I was almost going'  
           /a:/ /'pɛsə 'ɛrvaf/                      [pɛ 'sɛrvaf]                      'ask for herbs'  
           /u:/ /'ɪstu 'ɛrə a'sɪ/                      [ɪst'ɛrə 'sɪ]                      'it would be like this'

N.B. The deleted vowels are underlined.

Many varieties of English have the same features of strong and weak stress, with the stress positively correlated with vowel duration (Fry, 1955, Lieberman, 1960). In English, unstressed vowels tend to become lax and ultimately schwa. In addition, throughout the development of English, the final unstressed vowel has weakened to schwa and then ultimately dropped (Hyman, 1970). As explained in Section 2.2.3, not only the final vowel but also other vowels may be reduced to schwa and eventually deleted in certain phonological environments. Schwa in English may also be devoiced in pretonic position and in an extreme case it is deleted, which may be at the ultimate stage of vowel weakening as shown in (11).

- (11) *potato* /pə'teɪtəʊ/                      [pə'teɪtəʊ]                      -->                      ['pteɪtəʊ]  
       *police* /pə'li:s/                      [pə'li:s]                      --->                      ['pɪli:s]  
       *support* /sə'pɔ:t/                      [sə'pʰɔ:t]                      --->                      ['spʰɔ:t]  
       *director* /də'rektə/                      [də'rektə]                      -->                      ['dɹektə]  
       *ferocity* /fə'rɒsɪtɪ/                      [fə'rɒsɪtɪ]                      --->                      ['fɹɒsɪtɪ]

For all examples in (11), the pretonic /ə/ in these words is often not phonetically present in natural speech but may be present in citation form. The sequence /#pt-/ in *potato*, for example, violates English phonotactics. In the cases of *director* and *ferocity*, the liquids can be coarticulated with /d/ or /f/ respectively; /p/ of *support* may not be aspirated; /l/ of *police* may be partially or fully devoiced. Some speakers pronounce [ˈpɪli:s], [ˈdʒektə] as even citation forms for *police* and *director*. In these cases, it may possible to say that the schwa is deleted.

If there is even a slight influence of disappeared schwa, it would be better to say that the schwa is in fact present at least underlyingly, or the schwa is devoiced. A vowel may be devoiced, may not be present in terms of oral articulation but may retain its status keeping the syllabicity by making a following consonant syllabic, or may be deleted completely, reducing the number of syllables of a word. For example, schwa can optionally be deleted in the environment between a consonant and liquid which is followed by a weak vowel (12a). Although a following weak vowel is favoured for schwa deletion, it also occurs when schwa is followed by a stressed syllable (12b).

(12)a. [ə] ---> ø / C \_\_\_\_ [liquid] [-stress]

[ə]	--->	[ɹ]:	<i>preference</i>	[ˈpɹɛfəɹəns]	(3)	--->	[ˈpɹɛfrəns]	(2)
			<i>federal</i>	[ˈfɛdəɹəl]	(3)	--->	[ˈfɛdɹəl]	(2)
[əl]	--->	[l]:	<i>especially</i>	[ɪˈspeʃəli]	(4)	--->	[ɪˈspeʃli]	(3)
			<i>fatalist</i>	[ˈfeɪtəlɪst]	(3)	--->	[ˈfeɪtɪlɪst]	(2)

(12)b. [ə] ---> ø / C \_\_\_\_ [liquid] [+stress]

[ə]	--->	[ɹ]:	<i>correct</i>	[kəˈɹɛkt]	(2)	--->	[ˈkɹɛkt]	(1)
[əl]	--->	[l]:	<i>collect</i>	[kəˈlekt]	(2)	--->	[ˈkɹɛkt]	(1)

N.B. The figures in brackets are the number of syllables.



### 2.3.2.3 Speaking tempo and style

Many studies have reported that vowel devoicing occurs more at a faster speaking tempo and in casual speech (Johns, 1972, Semiloff-Zelasko, 1973, Dauer, 1980b, Gimson, 1980, Kohler, 1990, etc.). Since devoiced vowels are not phonemes in any reported languages, vowel devoicing is often treated as an optional fast speech or casual speech phenomenon in most languages.

Johns (1972: 65) studied vowel deletion in Continental Portuguese, and concluded that speaking tempo was one of the important factors in Continental Portuguese vowel devoicing. Adopting the classification of tempo for Mexican Spanish proposed by Harris (1969), Johns classified speaking tempo for Continental Portuguese as:

- (a) Largo: very slow, over precise
- (b) Andante: moderately slow, careful but natural
- (c) Allegretto: moderately fast, casual and colloquial
- (d) Presto: very fast, completely unguarded

Johns' conclusion was that devoicing probably occurred most extensively in the 'presto' style and was least likely to occur in the 'largo' style.

Dauer (1980b) studied vowel reduction in Greek and found that there seemed to be some correlation between speaking rate and vowel reduction, but speech style was more important. In a careful style, a speaker devoiced vowels, whereas in a casual style, more deletion of high vowels occurred and there was a tendency to undergo voicing assimilation. It was also observed that in careful style devoiced vowels were easily identified, while the segment boundary was not very clear in casual style. As far as speaking tempo was concerned, the faster the tempo, the greater was the frequency of vowel reduction. In words where a vowel was adjacent to or surrounded by voiceless vowels, not only high vowels but also /e/ and /o/ could be deleted in fast speech.

In modern English, vowel weakening is strongly related to stress, speaking rate and contextual influence. As mentioned earlier, as speaking rate increases, all vowels in unstressed syllables tend towards a common schwa (Lindblom, 1963: 1773). Similarly, as the degree of stress decreases, vowels seem to move towards a neutral or central vowel quality. The weakened vowels may be completely deleted in unaccented syllables in rapid colloquial speech. With a few exceptions such as *subjectivity* /sʌbdʒek'tɪvəti/, the further vowels are from the tonic syllable, the more the vowels in unstressed syllables tend to be /ə/ or /ɪ/. This process occurs mainly in words in common use, where vowels in a non-tonic syllable are likely to be reduced to /ə/, /ɪ/ or /ʊ/, unless the syllable carries a secondary stress.

#### 2.3.2.4 Sociolinguistic factors

Although voiceless vowels are not normally contrastive, they have significant sociolinguistic implications in some languages. For example, word-final vowel devoicing occurs in Yana women's speech; a vowel may be devoiced after a consonant and if the word-final vowel is devoiced, the preceding voiced consonant must also become voiceless. However, this phenomenon does not occur in Yana male speech (Donegan, 1985: 47). There seems to be a generation difference among Toposa speakers. Older people retain a full set of voiceless vowels, while younger people tend to pronounce and recognise fewer low- and mid-voiceless vowels (Schröder and Schröder, 1973: 18). This means that there is a gradual loss of voiceless vowels from low to high vowels.

In the generativist approach, vowel devoicing and deletion in (11) and (12) are optional, but Labov quoted by Zwicky (1970), proposed that vowel deletion was largely dependent on social class, speech style, speech rate, and so on.

### 2.3.2.5 Lexical, syntactic and semantic constraints

There are lexical and syntactic constraints on vowel devoicing. The same vowel in the same phonological environment in the same language may be devoiced or deleted in one word and not in another. Speakers of any language tend to articulate more precisely and pay more attention to difficult words and unfamiliar words, and as a result, vowel devoicing and deletion are less likely to occur.

Dauer's study (1980b) of Greek vowel devoicing found that high vowel devoicing and deletion occur most frequently in everyday words, such as common verb endings, articles and pronouns. However, devoicing is less common in formal words or uncommon words such as /fisi'kos/ 'physicist' and /ði'tia/ 'duty'.

Semiloff-Zelasko (1973) reported schwa deletion in semantically unfocussed words in Hebrew. A word initial unstressed schwa, which is derived from vowel neutralisation, in a semantically unfocussed word may be deleted in a small number of words with a very high frequency, such as /ħaxʃáv/: [əxʃáv] ---> [xʃav] 'now', /ħavál/: [əvál] ---> [val] 'but', /ħefʃár/: [əfʃák] ---> [(f)ʃak] 'it is possible'<sup>11</sup>. Although vowel neutralisation to [ə] occurs in unaccented word initial position, schwa deletion does not normally occur in content words and occur less frequently, such as /ħesríim/ [əskím] 'twenty', /ħasúr/ [əsúʔ] 'prohibited' where the neutralised [ə] is normally retained.

In English, schwa deletion mentioned in Section 2.3.2.4 may be applied to common words such as *nursery* ['nɜ:səɹi] ~ ['nɜ:sɹi], *factory* ['fæktəɹi] ~ ['fæktɹi], but schwa in the same environment is very likely to be retained in uncommon or difficult words such as *cursor* ['kɜ:səɹi], *livery* ['lɪvəɹi], *slavery* ['sleɪvəɹi] even in fast speech or casual speech.

---

<sup>11</sup> Although Semiloff-Zelasko did not mention it in her paper, word initial glottal stop /ʔ/ and voiced pharyngeal fricative /ɣ/ exist in phonemic transcription in Hebrew, but seem to be normally deleted in natural speech.





reduced to a single consonant, which is a general process in German (see next section). However, the process has syntactic and semantic constraints, in other words the process is blocked if the reduction causes a semantic confusion. Kohler (1990: 75) presented examples of the numeral ['axtnn'oıntsɪç] '98' always with a double [n] in order to differentiate it from the short form for the money amount '8 Mark 90 Pfennig' which is '8.90' ['axtn'oıntsɪç] with a single [n].

Kohler concluded that "when syntactic and semantic redundancies do not weight one word class more highly than the other and when a differentiation is essential for effective speech communication, the reductions stop at points where the phonetic forms are still distinct" (Kohler, 1990: 75).

## 2.4 Processes of vowel reduction

Vowel devoicing therefore in many languages occurs as part of the process of vowel weakening and it seems to be closely related to speaking rate and speech style. The reduction process is optional, and the realisation of sounds can be at any stage from fully voiced non-reduced form to complete deletion (see Section 2.2.2). Many studies on various languages have reported this gradual vowel reduction process. As mentioned earlier (Section 2.3.2.2), the vowel reduction process in many languages often involves vowel centralisation, neutralisation, devoicing and deletion, but different processes operate in each language. In this section, I will examine vowel reduction processes in Greek, Hebrew, and German, and will also consider how these processes occur citing examples in Korean.

### 2.4.1 Greek vowel reduction processes

Stress plays an important role in Modern Greek, and syllables do not have the same strength. Vowel duration is dependent on the presence or absence of stress: stressed vowels have longer duration than unstressed vowels (Dauer, 1980a: 84). The vowel

/i/ is very common as an unstressed vowel and therefore it is often deleted or devoiced in certain contexts. The vowel /u/, which is the least common vowel in Greek, also undergoes the same processes. Vowel reduction in Greek is treated as an optional conversational fast speech process, but it is controlled by its phonetic environment, position of stress, speech style and so on. Devoicing of high vowels occurs most frequently in a post-stressed syllable. This phenomenon is very common in the northern dialects where unstressed /e/ and /o/ are often realised as [i] and [u] both of which are frequently deleted (Newton, 1972: 182).

Dauer (1980b) considered that devoiced vowels in Greek occur during the vowel reduction process and proposed different degrees of vowel reduction: (i) when the vowel is fully voiced, (ii) when the vowel is short, (iii) presence of voicing of brief duration, (iv) no voicing with the release of a preceding consonant, and (v) no vowel. Factors mentioned in Section 2.3.2 influence vowel reduction. These five stages are applicable to unstressed high vowels, and are affected by various phonetic factors. Dauer's description of the five stages of Greek vowel reduction is as follows.

(i) When the vowel is full, voiced, and has a formant structure similar to when it is stressed. Although unstressed /i/ and /u/ generally have lower intensity and are shorter in duration than when they are stressed, their quality is very clear. This reduction takes place when the vowels are preceded and followed by voiced consonants.

(ii) This reduction occurs after a liquid or nasal. Although the vowel has a complete formant structure, it is very difficult to segment from a preceding sound. In other words, these vowels are auditorily perceived as syllabic nasals or laterals, or as slightly lengthened nasals or laterals.

(iii) The voicing of the vowel is still present but an irregular waveform with low intensity is detected. The duration is very brief. The vowels are very often perceived as very short centralised vowels. This usually occurs between voiceless and voiced consonants, but sometimes it occurs between voiceless stops and fricatives.

(iv) This occurs between voiceless consonants, or after voiceless consonants at the end of a phrase with a falling intonation pattern. There is no voicing, but a short period of friction or voicelessness, which produces the sound as /i/ or /u/. They are in fact

completely devoiced whispered [i] and [u], which have the same duration as short /i/ or /u/.

(v) This mostly occurs between voiceless consonants, but sometimes occurs even when one of the consonants is voiced. The preceding consonant is released directly into the following consonant so that no separate vowel is heard. There is no trace of the vowel. It is, however, possible to detect the underlying vowel because of the palatalisation or non-palatalisation of the preceding consonant, which leads native Greek speakers to sense the underlying vowel. The difference between (iv) and (v) is very slight. It is a matter of timing.

Dauer's description of Greek vowel reduction is also applicable to other languages, especially those in which the function of stress is very important.

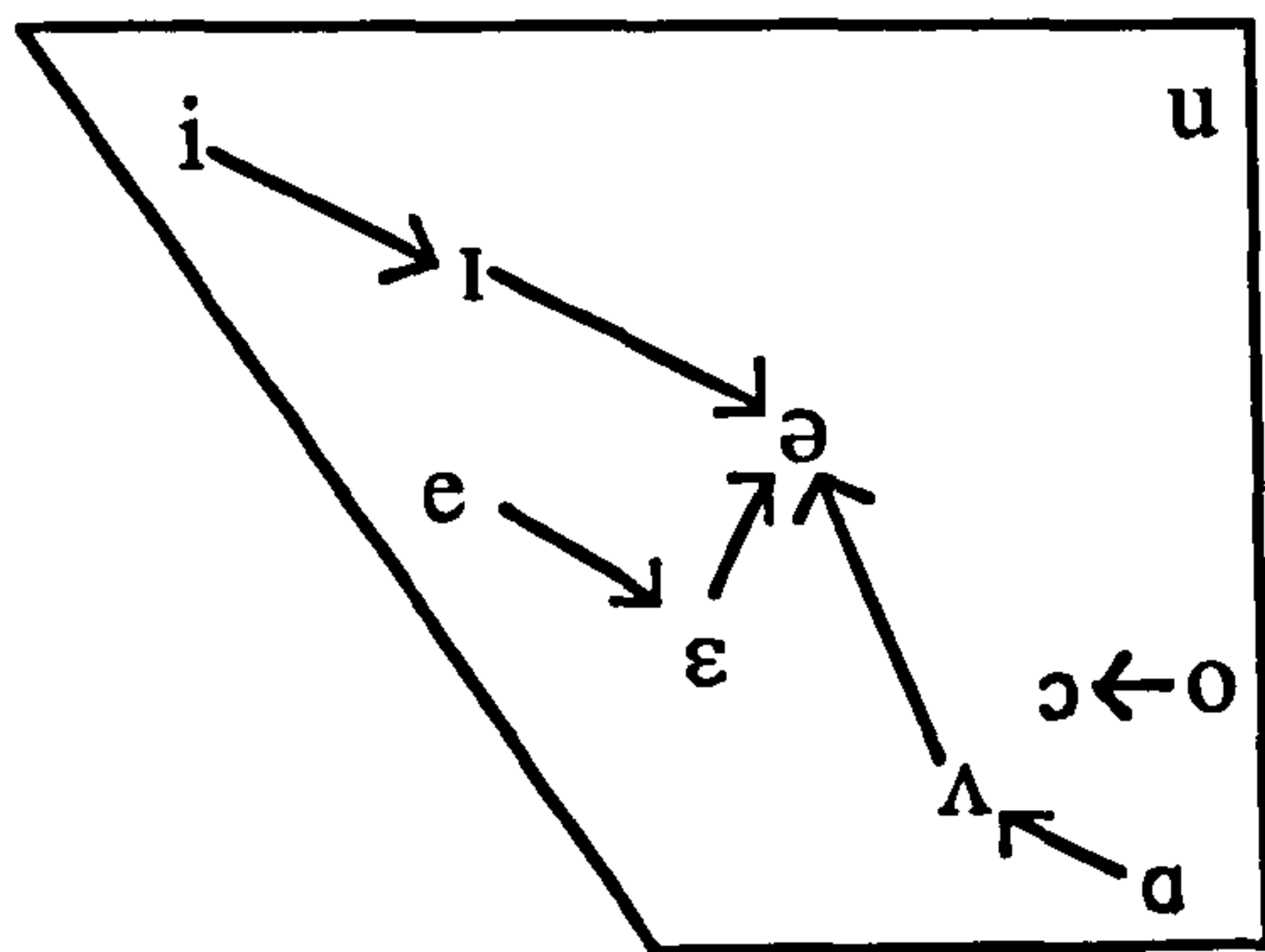
#### 2.4.2 Hebrew vowel reduction processes

Similar processes to Greek vowel reduction can be seen in Hebrew examples. Hebrew has five vowels /i, e, a, o, u/, and the non-rounded vowels /i, e, a/ undergo reduction processes in fast speech, and then they are eventually deleted. There are six processes for vowel reduction: (i) vowel laxing, (ii) vowel neutralisation, (iii) monophthongisation of resonants, (iv) vowel frication, (v) degemination and (vi) desyllabification (Semiloff-Zelasko, 1973).

Firstly, the vowels except for /u/ become lax especially in unstressed position in fast speech, then non-rounded vowels /i, e, a/ are neutralised to [ə]. Processes (i) and (ii) are shown in Figure 2.1.

(15)	/i/:	[i]	--->	[ɪ]
	/e/:	[e]	--->	[ɛ]
	/a/:	[a]	--->	[ʌ]
	/o/:	[o]	--->	[ɔ]

Figure 2.1 Laxing and neutralisation of Hebrew vowels



(Semiloff-Zelasko, 1973: 54)

Process (iii) may follow after processes (i) and (ii) . A vowel takes the colour of the following resonant without loosing its syllabicity, then the resonant sequence monophthongises to a single resonance, depending on speech style and tempo as shown in (16).

(16) əR	--->	R̥R	--->	R̥ <sup>12</sup>	
/ʃɛl/	--->	[ʃəl]	--->	[ʃl̥]	---> [ʃl̥] 'of'
/gam/	--->	[gəm]	--->	[gm̥]	---> [gm̥] 'also'
/dvaʁím/	--->	[dvəʁím]	--->	[dvəʁím̥]	---> [dvəʁím̥] 'things'

Process (v) is ‘vowel frication’, which is a terminology used by Semiloff-Zelasko (1973), meaning that, as in process (iv), a neutralised vowel loses its resonance between fricatives without loosing its syllabicity as shown in (17).

(17)/ʃiʃím/	[ʃiʃím]	‘sixty’
	↓	
	[ʃɪʃím]	laxing
	↓	
	[ʃɪʃím̥]	devoicing
	↓	
	[ʃʃʃím̥]	vowel frication
	↓	
	[ʃʃím̥]	degemination

<sup>12</sup> Although the process is preceded by vowel neutralisation, only examples of /e/ and /a/ are observed in practice, because neutralisation of /i/ to [ə] is relatively infrequent.



When speaking rate increases, the vowels through processes (iv) and (v) will eventually be deleted as in (18).

(18)/dérex/	[déɤɤx]	'way'
	↓	
	[déɤəx]	neutralisation
	↓	
	[déɤ̥əx]	devoicing
	↓	
	[déɤ̥ɤx]	monophongisation
	↓	
	[déɤ̥ɤx]	desyllabification
	↓	
	[déɤ̥x]	degemination (vowel deletion)

(both (17) and (18) are taken from Semiloff-Zelasko, 1973: 57)

Since prefixes and suffixes are almost always unstressed, vowels in prefixes and suffixes undergo neutralisation and deletion processes. For instance, suffixed /-ɛt/ (female singular), /-ʔɛt/ (direct object) and /ʔat/ ('you' female singular) can all be realised as [t] possibly via [\*ət]<sup>13</sup>. For example;

(19) /noténɛt/	--->	[notént]	'given (female)'
/ʔoméɛt/	--->	[ʔomért]	'says (female)'

Similarly, a vowel in a prefix can be deleted. When it occurs, the consonant in the prefix which became syllable initial may become a syllable associated with the preceding word as long as the new cluster does not violate Hebrew phonotactics. (19) shows examples of resyllabification and vowel deletion in (a) /hi malamédɛt/ 'she teaches' and (b) /smona bastej/ 'eight in two'.

(20)	a. /hi malamédɛt/	b. /smona bastej/
resyllabification	[him əlamédɛt]	[smonab astej]
vowel deletion	[him lamédɛt]	[smonab stej]

<sup>13</sup> "\*" means non-occurring form.

### 2.4.3 German vowel reduction processes

Similar processes of vowel neutralisation and reduction are observed in German. As explained in Section 2.3.2.5, German vowels are reduced in unstressed position, which means durational reduction, opening, centralisation, monophthongisation of vowels, and vowel and consonant deletion. As a result of the reduction process, the same word may have several phonetic realisations forming a reduction hierarchy which correlates with the lowering of the stylistic level and with the degree of familiarity in the communicative situation (Kohler, 1990: 72). Vowels are centralised and neutralised to schwa [ə], and then may be deleted. Vowel deletion creates consonant clusters, and geminate consonants, both of which may also be reduced; geminate consonants may be reduced to a single consonant, aspiration of a preceding consonant of consonant clusters may be lost, and the place, manner or voicing of the consonants may be assimilated to each other.

(21) a. <i>mit dem Wagen</i>	'by car'
[mɪtʰ de:m v'a:gən]	Very careful pronunciation
[mɪtʰ ɖe:m v'a:gən]	Devoicing of voiced stop
[mɪtʰ ɖem v'a:gən]	Vowel reduction in function word
[mɪtʰ ɖm v'a:gn]	[ə] deletion before nasal consonant
[mɪt ɖm v'a:gn]	Deletion of aspiration before stop
[mɪp ɓm v'a:gn]	Regressive place assimilation
[mɪp ɓm v'a:gŋ]	Progressive manner assimilation
[mɪ ɓm v'a:gŋ]	Reduction of double consonants
[mɪ bm v'a:gŋ]	Voicing of plosive in unstressed function word
[mɪ mm v'a:ŋŋ]	Nasal assimilation
[mɪ m v'a:ŋ]	Reduction of double consonants

(21) b. <i>Er hat mir geholfen.</i>	'He has helped me.'
[ʔe:ɤ hat <sup>h</sup> mi:ɤ gəh'ɔlfən]	Very careful pronunciation
[e:ɤ hat <sup>h</sup> mi:ɤ gəh'ɔlfən]	Deletion of glottal stop in function word
[ɛɤ hət <sup>h</sup> məɤ gəh'ɔlfən]	Vowel reduction in function word
[ɛɤ hət <sup>h</sup> məɤ gəh'ɔlfn]	[ə] deletion before nasal consonant
[ɛɤ hət <sup>h</sup> mə gəh'ɔlfn]	Word final [ɤ] vocalisation
[ɛɤ hət <sup>h</sup> mə gəh'ɔlfn]	Word final [ɤ] vocalisation
[ɛɤ hət mə gəh'ɔlfn]	Deletion of aspiration before stop
[ɛɤ ət mə gəh'ɔlfn]	[h] deletion in function word
[ɛɤ t mə gəh'ɔlfn]	Word initial vowel deletion in function word after /r/ vocalisation in the preceding word
[ɛɤ p mə gəh'ɔlfn]	Regressive place assimilation
[ɛɤ b mə gəh'ɔlfn]	Voicing of plosive in unstressed function word

(21a) and (21b) show a list of various but by no means all possible phonetic realisations of the same lexical items undergone reduction processes depending on the speaking rate and style (Kohler, 1990: 82-83).

Reduction of vowels and consonants is typically common when they are adjacent to nasals, or preceded by plosives and followed by nasals. Kohler considered speech rhythm and articulatory gestures as factors which influence the reduction processes. First of all, [ə] shortening as the consequence of the general shortening in unstressed position is common because of the isochrony of the interval between stressed syllables. Although the isochrony is not perfect, there is a general tendency in German for function words and unstressed syllables to be compressed durationally.

Secondly, articulatory gesture plays an important role regarding the deletion of [ə] and aspiration and assimilation of place, because of an articulatory restructuring due to a minimisation of energy expenditure which limits the extent of movements and reduces the number of moving organs involved in speech production. All these above processes are typically common either between nasals or between stop and nasal, and the places of articulation of the preceding and following consonants involved are /labial



+ apical/, /apical + labial/, /dorsal + apical/, /dorsal + labial/, /apical + apical/ or /labial + labial/. As far as the type of preceding and following consonants is concerned, the articulatory movements of the following nasal partially or completely overlap with the preceding consonantal gesture. When the duration of [ə] is extremely short or the [ə] is completely deleted, velic opening and oral occlusion for the following nasal are so much advanced in time that they are established at the offset of the preceding consonant. Kohler explained this process from an articulatory point of view:

“Tongue dorsum and lip gestures, in close connection with jaw movements, constitute large, global processes, associated with, for example the large extrinsic tongue muscles. Tongue apex articulations require much more finely controlled and more precisely tuned, and therefore also more costly, adjustments, linked with the smaller intrinsic tongue muscles. When, for perceptual reasons, the fine control is not absolutely necessary and can thus be neglected without harm to the communicative goal, it may be dispensed with under certain situational conditions such as higher speech rate or relaxed speech, when there is less time for finer control or when the energy expenditure per unit time is reduced to the minimum necessary for a successful speech act.” (Kohler, 1990: 88).

#### 2.4.4 Gestural overlap analysis of vowel devoicing in Korean

Jun (1993) and Jun and Beckman (1993a, 1993b, 1994) discussed vowel devoicing in Korean with relation to the gestural overlap which is a theory proposed by Browman and Goldstein (1989, 1990a, 1990b). The gestural overlap theory explains the existence of a segment which otherwise seems to be deleted. Taking an example of an English casual pronunciation *perfect memory* [p<sup>h</sup>əfəkməməɹi] in which the last consonant /t/ of /p<sup>h</sup>əfəkt/ is not heard, Browman and Goldstein explained that articulatory data actually showed clear evidence of the lingual gesture for the alveolar consonant /t/. The alveolar gesture was hidden because it overlapped the velar and bilabial gestures of the preceding and following consonants [k] and [m]. Therefore, the /t/ appeared to be deleted.

In Korean, high vowels /i, u, ʉ/ are devoiced between voiceless consonants. Korean has three different kinds of stops, namely aspirated stops, fortis stops and lenis stops (see Section 2.2.2). These stops show different glottal openings and vowel devoicing



rates were largely dependent on the type of preceding stops: the wider the glottal width, the greater is the degree of gestural overlap between voiceless stops and vowels, i.e. vowels are more devoiced. Kagaya (1974) and D.-W. Kim et al. (1992) showed that an aspirated stop has a large glottal-opening gesture that is at its peak opening around the oral release, and that a fortis stop has a much smaller opening that ends in a tightly closed glottis well before the oral release. The glottal movement of a lenis stop shows two different patterns depending on the position where it occurs: (a) in initial position, it has a fairly large glottal opening like the aspirated stop, but timed so that the glottis is nearly closed again when the vowel starts, and (b) in medial position, there is often no apparent opening between the vocal folds.

The results of Jun and Beckman (1993 and 1994) and Jun (1993a) found that both complete and partial vowel devoicing was most prevalent when vowels were preceded by aspirated stops which showed the largest glottal opening gesture. When the vowels occurred after fortis stops, the devoicing rate was the lowest for either complete or partial devoicing, since fortis stops have a much smaller glottal opening. For the vowels preceded by lenis stops, two patterns of devoicing rates were obtained, i.e. in phrase initial position and phrase medial position. The devoicing rate of vowels preceded by lenis stops in phrase initial position was much higher than in phrase medial position as predicted by the different glottal opening gestures for lenis stops in those positions. In initial position, the glottal opening gesture is large to start with and gets narrower gradually, then closed just after an oral release. In medial position, on the other hand, the glottis is tightly closed throughout its articulation.

Jun (1993a) also examined the effect of speaking tempo on vowel devoicing rates. She found that vowel devoicing rates increased at slow tempo and decreased at fast tempo. This was because the number of phrases decreased at fast tempo, the number of phrase initial lenis stop also decreased at fast tempo, and so did the number of clearly voiceless lenis stops. She also proposed another reason for lower devoicing rates at fast tempo that although the lenis stop is almost always voiceless at Accentual Phrase initial position, it was sometimes voiced in the same position at the faster rate (Jun, 1993a: 102).

Jun and Beckman's gestural overlap analysis may also be applicable to the vowel reduction processes in other languages. It is because (a) vowels undergoing reduction processes are predominantly high vowels and schwa all of which are intrinsically short in duration, (b) typical environments for vowel reduction are adjacent to voiceless consonants or a pause for both of which the glottis is wide open, (c) another environment favoured for vowel reduction is adjacent to nasals, for which articulatory gestures start so early that overlap with an adjacent short vowel, and (d) vowel reduction processes are usually more prevalent in fast and casual speech, in both of which the movements of the articulators are likely to be less well executed.

## 2.5 Conclusions

In this chapter, I have defined 'vowel devoicing', and examined it and related phenomena in various languages. In theory, vowel devoicing is a change of phonation from voiced to voiceless. However, in reality, vowel devoicing is closely related to vowel reduction and deletion, and it is often part of the vowel reduction process. Vowel devoicing also has to be considered in relation to the phonology of the language in question, including phonotactics, accentuation, and syllable structure. This is because the presence of the 'devoiced' vowel is retained in features such as syllabicity, presence of accent on a devoiced vowel, and duration.

The most likely candidates for vowel devoicing in most languages are high vowels and schwa, all of which are inherently short in duration, especially when they are adjacent to a voiceless consonant. Devoicing typically occurs in unstressed position, most commonly in fast and casual speech, where vowels are centralised, shortened, and eventually devoiced. Devoicing is also common in function words and common words, and less likely to occur in rare words, semantically important or unfamiliar words. Voiceless vowels do not have phonemic status in any language, and are always considered as allophones of their voiced counterparts.

In the later chapters, I will examine Japanese vowel devoicing processes in particular the differences and similarities to the processes in other languages.

# Chapter 3

## Japanese phonology

### 3.1 Introduction

Before I discuss Japanese vowel devoicing, I will briefly describe some aspects of Japanese phonology. Within Japanese phonology there are many important aspects, but I will focus on the areas which are relevant to vowel devoicing and are helpful to read this thesis, i.e. (1) syllables and morae, (2) accent and intonation, (3) vowels and (4) consonants.

There are many dialects in Japan. Each dialect has its own phonological system and rules, many of them having very different phonological systems and rules from Standard Japanese. However, in this chapter, I will mainly discuss the phonology of Standard Japanese (“Hyoojun-go”) which is spoken in the Tokyo area, unless otherwise stated, because vowel devoicing is a typical characteristic of mainly eastern Japanese dialects, including Standard Japanese.

### 3.2 Japanese syllable structure and mora

In order to examine Japanese vowel devoicing, it is necessary to understand Japanese syllable structure and morae. Both the syllable and the mora are important units in Japanese. The main emphasis of studies on Japanese syllables and morae have been durational aspects, with the mora being considered as a basic phonological unit of Japanese. These durational aspects of the mora and syllable will be discussed in



Chapter 6. This section will focus on other phonological characteristics of Japanese morae and syllables.

The vowel is the minimum central element for both the syllable and the mora, except moraic nasals and moraic consonants (see below). When a vowel is devoiced, there is no voiced vowel element in the syllable and mora, and they are acoustically a sequence of voiceless consonants. However, the devoiced vowel normally keeps its vowel status as a 'devoiced vowel' within the syllable or mora, similar to languages such as Mandarin Chinese and Toposa discussed in Chapter 2, because deleting a vowel creates various consonant clusters which are not favoured in Japanese syllable structure.

The basic Japanese syllable structure is CV, but various syllable structures are possible within the pattern shown in Figure 3.1. A minimum constituent of a syllable is V. This syllable structure is also imposed on foreign loan words: a vowel is inserted between consonants, and after a consonant in cases of consonant clusters and syllable final consonants. For instance, a loan word from English *text* /tekst/ is phonemicised as /tekusuto/ in Japanese, so that the sequences do not violate Japanese syllable structure.

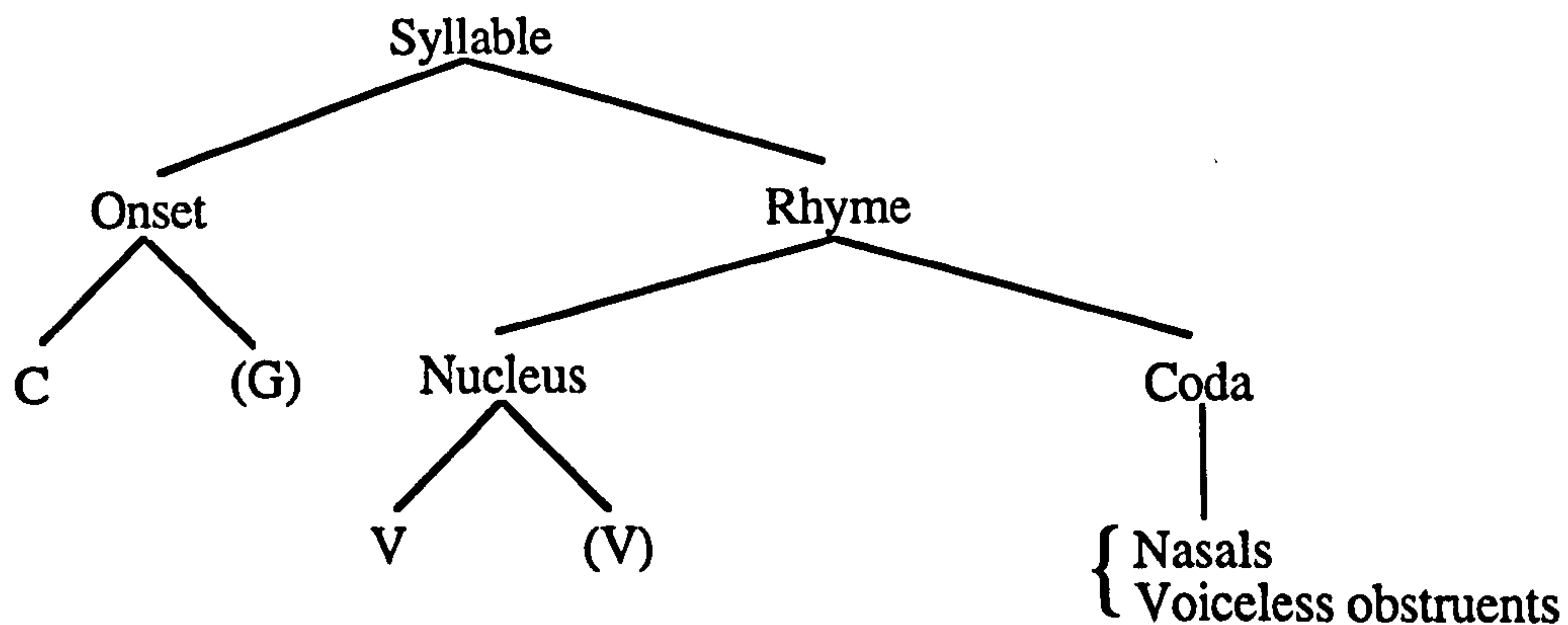
(1a) - (1d) in the next page are light syllables and (1e) and (1f) are heavy syllables. Any consonant can be found in the syllable initial position, but the glide in second position is normally /j/.<sup>1</sup> The place of articulation of nasals and voiceless obstruents in syllable final position, except when followed by a vowel or in word final position, is predictable from the following consonants. The nasal in syllable final position is a 'moraic nasal' (see below in this section and Section 3.5.2), and a syllable final voiceless obstruent which is the initial part of a geminate consonant also constitutes a mora on its own (see Section 3.5.3).

---

<sup>1</sup> The glide in second position can be /w/ only in some dialects, e.g. /kwaji/ 'fire', /kwaNkei/ 'relationship'.

**Figure 3.1 Japanese syllable structure<sup>2</sup> (Abe, 1986)**

(Here "G" means a glide.)



(1a) to (1f) list permitted syllable structures in Japanese.

(1) Possible syllable structures:  $(C)(j)V(V)\left(\begin{Bmatrix} N \\ C \end{Bmatrix}\right)$

- |    |             |                      |                    |                                |
|----|-------------|----------------------|--------------------|--------------------------------|
| a. | $(C)(j)V$   | : /i/ 'stomach'      | /ki/ 'tree'        | /tja/ 'tea'                    |
| b. | $(C)(j)VV$  | : /ai/ 'love'        | /koi/ 'carp'       | /gjoo/ 'line'                  |
| c. | $(C)(j)VN$  | : /uN/ 'luck'        | /kaN/ 'can'        | /sjuNtoo/ 'general strike'     |
| d. | $(C)(j)VC$  | : /utta/ 'sold'      | /hotta/ 'dug'      | /sjotta/ 'put on the shoulder' |
| e. | $(C)(j)VVN$ | : /saiN/ 'signature' | /siiN/ 'scene'     | /tjeeN/ 'chain'                |
| f. | $(C)(j)VVC$ | : /kootte/ 'frozen'  | /kaette/ 'instead' | /gjuutto/ 'tightly'            |

<sup>2</sup> This is the most widely acknowledged syllable structure of Japanese, but by no means the only one. For instance, Katada (1990) complains the fact that this syllable model does not represent a unit  $(C)(j)V$  ('mora'), another important unit in Japanese. Kubozono (1989) also questions whether a sub-unit of a syllable 'rhyme' has any role to play in 'mora' based languages such as Japanese. Therefore, he proposed an alternative syllable structure (4) below in this section. Yoshida (1990) based on the framework of government phonology also argues that this model cannot deal with a sequence  $(C)(j)VNC$  such as /roNdoNkko/ 'Londoner', /hoNtte/ '(someone says that) books', and that sequences  $(C)VV$  and  $(C)VN$  cannot constitute a single syllable but are bisyllabic. He rejects the concept of mora in Japanese, and proposes that the second V in Nucleus and N in Coda constitute a syllable on their own. Tabata (1989) claims that a multi-tier analysis is needed in order to account for a sequence such as  $(C)(j)VNC$ . In this section, however, I will discuss syllable structure based on Abe's model.

The syllable is not the smallest rhythmic unit in Japanese. The mora has traditionally been considered as a basic unit in Japanese. A basic mora unit is (C)V, although both nasals and voiceless obstruents in syllable final position can constitute morae on their own (see Section 3.5.2 and 3.5.3).

The difference between syllable and mora is shown in the words listed in (2).

(2)		<u>syllable</u>	<u>mora</u>
a.	/ko.-to/ 'harp'	2	2
b.	/ko.o/ 'the instep'	1	2
c.	/ko.N/ 'dark blue'	1	2
d.	/ka.N.-da/ 'place name'	2	3
e.	/ka.t.-ta/ 'buy (past)'	2	3

N.B. Here and hereafter,

Syllable boundaries are marked by "-".

Mora boundaries are marked by ".".

(Sugito, 1989: 161)

As mentioned earlier in this section, the mora is defined as a phonological rhythmic unit. It has been said that each mora takes roughly the same length of time to say (Jōo, 1977: 115, Ladefoged, 1982: 226). However, recent acoustic studies have revealed that there is no real isochrony of Japanese morae, although the results of the studies are controversial (Beckman, 1982: 133-134, Kawasaki, 1983: 393-396, Vance, 1987: 63, etc.). On the other hand, some studies support mora-based durational characteristics of Japanese (Sugito and Fujisaki, 1977, Port et al., 1982 and 1987, Sugito, 1989, Han, 1994, etc.). Durational aspects of morae in Japanese will be discussed in Chapter 6.

It has been argued that the mora is a psychological unit strongly influenced by orthography, because each letter in two types of Japanese writing systems (hiragana and katakana) represents a mora (Beckman, 1982, Vance, 1987: 72). Kubozono (1989) found some evidence that speech errors in Japanese occur not by a segmental

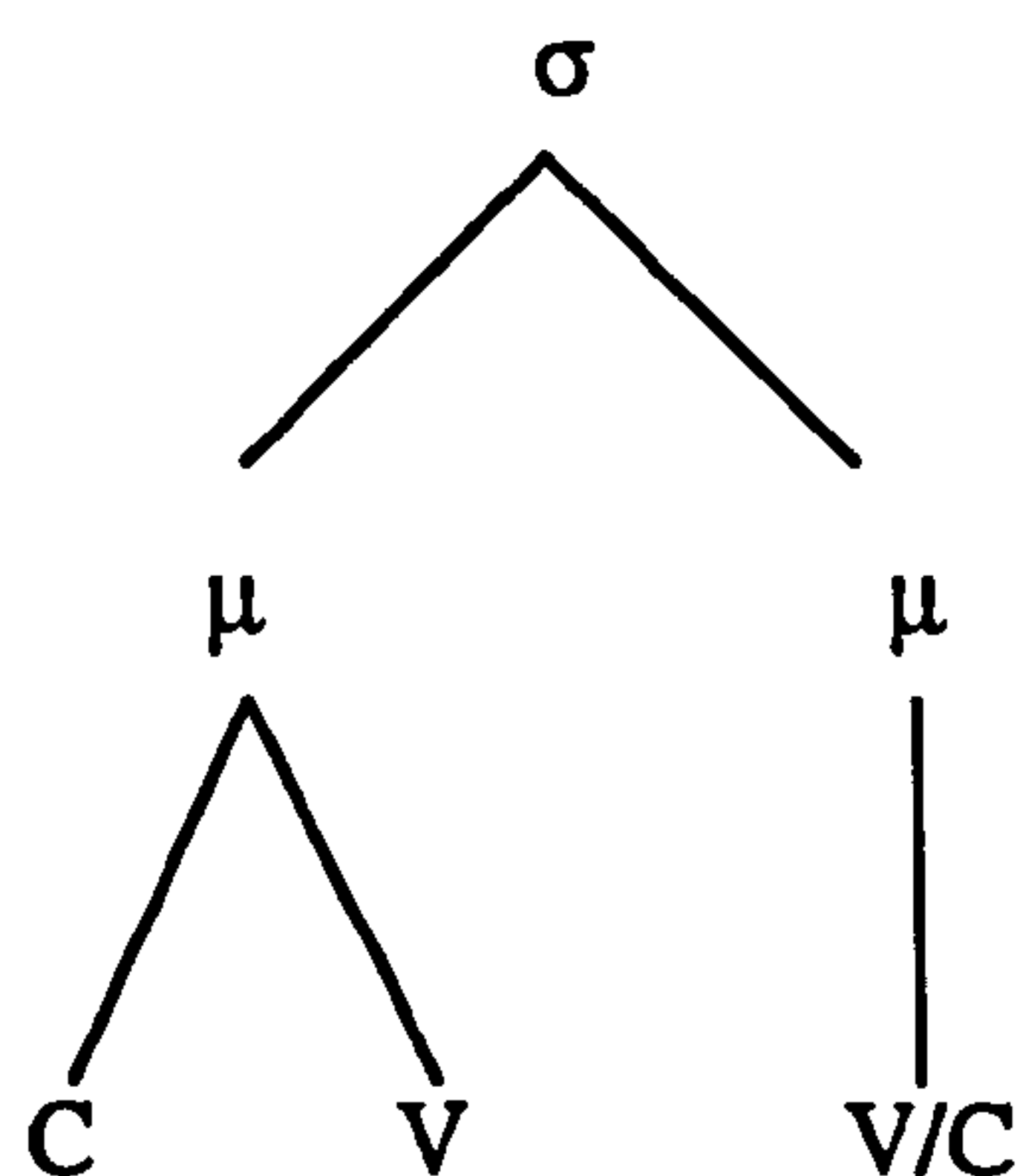
or syllabic unit but by a moraic unit. Some examples are listed in (3a) - (3d) (all examples are taken from Kubozono, 1989: 252).

- (3) a. mo.o-ta.a.-ba.i.-ku 'motorbike'  
       ---> mo.i.-ta.a.-ba.i.-ku  
       b. zju.u.-go.-pa.a-se.N.-to 'fifteen percent'  
       ---> zju.u.-go.-pa.N.-se.N.-to  
       c. ko.-ma.f.-te.-i.-ru 'being troubled'  
       ---> ko.-ma.N.-te.-i.-ru  
       d. ka.N.-ke.i.-ka.i.-ze.N 'relationship improvement'  
       ---> ka.i.-ke.N.-ka.i.-ze.N

Kubozono concluded that the moraic unit is essential for significant generalisations about linguistic phenomena in Japanese, whereas syllable boundaries did not seem very relevant.

Kubozono (1993) carried out word blending experiments, in which subjects were given auditory stimuli consisting of a pair of monosyllabic words and were asked to blend them into a monosyllabic form. He compared speakers of different languages and found that Japanese speakers used a perceptual segmentation strategy based on the mora (/CV/). The relationship between the mora and the syllable is as shown in (4). Within the syllable ( $\sigma$ ), the nuclear vowel is combined with the preceding consonant to form a mora ( $\mu$ ), whereas the post vocalic consonant or a post nuclear vowel (i.e. the second half of a long vowel or a diphthong) forms another mora as shown in (4).

(4)



“σ” denotes syllable.

“μ” denotes mora.



In the word blending experiments, a subject was asked to listen to two monosyllabic words and create a new monosyllabic word from the two. His results suggested that when Japanese speakers tended to blend words as CV + C rather than C + VC. For example, Japanese speakers created *fat* /fæt/ (CV/C) rather than *foot* /fut/ (C/VC) from two English words *fan* /fæn/ and *put* /put/, similarly they made *fight* /fat/ (CV/C) from *five* /faɪv/ and *sheet* /ʃi:t/ rather than *feet* /fi:t/ (C/VC) which English speakers did.

Katada (1990) demonstrated that even preliterate Japanese speaking children use the mora as a unit of Japanese, rather than the syllable, in a language game called "shiritori (hip-taking)", which means to take the last mora of a word and create a new word starting with the same mora of the last word. To play the game, one has to be able to identify the final mora of a word. For instance:

(5) tu.ba.me 'swallow'

me.da.ka 'killifish'

ka.o 'face'

o.N.ga.ku 'music'

:

The game carries on until someone says a word which ends with /N/, because no Japanese words start with /N/. Japanese speaking preliterate children can segment words by mora, thus they do not play the game in the manners of (6a) and (6b):

(6) a. go.-ha.N 'meal'

\*ha.N.-ta.i 'objection' (by syllabic unit)

b. ri.N-go 'apple'

\*o.ri.ga.mi 'origami' (by phonemic unit)





no reason why the second mora of a syllable could not bear an accent. Therefore, Japanese accentuation rules must refer to both moraic and syllabic units.

The Japanese accent system will be discussed in more detail in the next section.

### 3.3 Accent and intonation

It is essential to understand the patterns of Japanese accent and intonation, in order to study vowel devoicing. When an accented vowel is devoiced, it is still regarded as accented, though theoretically a listener cannot hear the pitch movement of the devoiced vowel. Standard Japanese is a pitch accent language. Every single word has a fixed pitch pattern. When there is pitch drop, the preceding mora is said to be accented. The pitch pattern is differentiated by the location of the pitch drop position, which can either be after any mora in the word or be absent, i.e. non-accented. The mora immediately before the pitch drop is said to be accented. The pitch after the drop is low. The pitch up to the drop is high, except for an initial mora which is always low unless it is the accented mora. No word can have more than one accented mora.

As shown in Table 3.1,  $n$ -mora words have  $n+1$  pitch patterns. The last pattern of each set is unaccented. The last two patterns of each set are differentiated by whether there is a pitch drop after the last mora or not, when a particle or auxiliary follows. (9) lists examples when a subject marker 'ga' follows the words.

Table 3.1 Pitch patterns in Standard Japanese

N.B."O" represents a mora.

“ ` ” indicates a location of pitch drop, i.e. a mora preceding “ ` ” is said to be accented.

	(a)	(b)	(c)	(d)	(e)
(i) 1-mora word:	O`	O			
(ii) 2-mora word:	O`O	OO`	OO		
(iii) 3-mora word:	O`OO	OO`O	OOO`	OOO	
(iv) 4-mora word:	O`OOO	OO`OO	OOO`O	OOOO`	OOOO
		:			
		:			

(9) Pitch patterns of words followed by a subject marker /ga/

	(a)	(b)	(c)	(d)	(e)
1. /hī`.ga/	/hi.gā/				
‘fire’	‘the sun’				
2. /hā`.si.ga/	/ha.sī`.ga/	/ha.sī.gā/			
‘chopsticks’	‘bridge’	‘edge’			
3. /kā`.ra.su.ga/	/ko.kō`.ro.ga/	/a.tā.mā`.ga/	/sa.kā.nā.gā/		
‘crow’	‘heart’	‘head’	‘fish’		
4. /kā`.ma.ki.ri.ga/	/mu.rā`.sa.ki.ga/	/ka.mī.nā`.ri.ga/	/o.tō.o.tō`.ga/	/ni.wā.tō.ri.gā/	
‘mantis’	‘purple’	‘thunder’	‘brother’	‘hen’	



The principal factor controlling lexical accent in Standard Japanese is not the assignment of high or low pitch to each mora, but rather the location of the accent nucleus<sup>3</sup>. The pitch rising point is specified in the patterns, falling between the low pitched first mora and the high pitched second mora of a word, unless the first mora is accented or another word precedes it in an accentual phrase.

The mora which immediately precedes the pitch drop is called the “accent nucleus” (Ueno, 1989: 185). The accent nucleus is usually marked by the symbol // in Japanese phonology. Acoustic studies have revealed that the actual pitch drop occurs slightly after the phonological accent nucleus (Ueno, 1977: 287, Fujisaki and Sugito, 1977: 101). However, native Japanese listeners agree on the accent nucleus placement, even when the vowel is voiceless. This will be discussed further in Chapter 4, Section 4.5.2.

It should be noted as mentioned in Section 3.2, that the moraic nasal /N/, the second element of a long vowel or a vowel sequence, and the first element of consonant geminations cannot bear the accent nucleus in Standard Japanese (Vance, 1987: 65-66, Shibatani, 1987: 140).

### 3.4 Vowels

Most of dialects<sup>4</sup> in Japanese including Standard Japanese have a five vowel system, namely /i, e, a, o, u/ as shown in Table 3.2 (Jōo, 1977: 121, Shibatani, 1987: 137).

---

<sup>3</sup> It is important to note that a phonological sequence from High to Low (HL), is phonetically a sharp drop of F0 from H to L (see Appendix I, Figures 1(a) and 1(b)), whereas that of LH is a very little gradual change of F0, phonetically it is almost a level tone (see Appendix I, Figure 1(c)). In other words, although phonologically the pitch pattern of a word may be HLL and that of another word may be LHH actual F0 contours of these two pitch patterns are not the reverse of each other. F0 change from H to L is always a very sharp fall, whereas the F0 change from L to H is very little. The F0 contours of the three pitch patterns, HLL, LHL and LHH are shown in Appendix I, Figures 1(a) ~ 1(c). In the sequence HLL, there is a sharp fall of F0 from H to L, and thereafter the fall is more gradual. In the sequence LHL, F0 shows a relatively steep rise from L to H, followed by a more sharp fall from H to L. In the sequence LHH, F0 rises only slightly from L to H, and thereafter remains fairly constant.

<sup>4</sup> Vowel systems in Japanese dialects range from a three-vowel system (/i, a, u/) in the Yonaguni dialect of Okinawa, to an eight-vowel system in the Nagoya dialect (the standard five vowels plus /ü, ö, æ/) (Jōo, 1977: 121, Shibatani, 1987: 137)

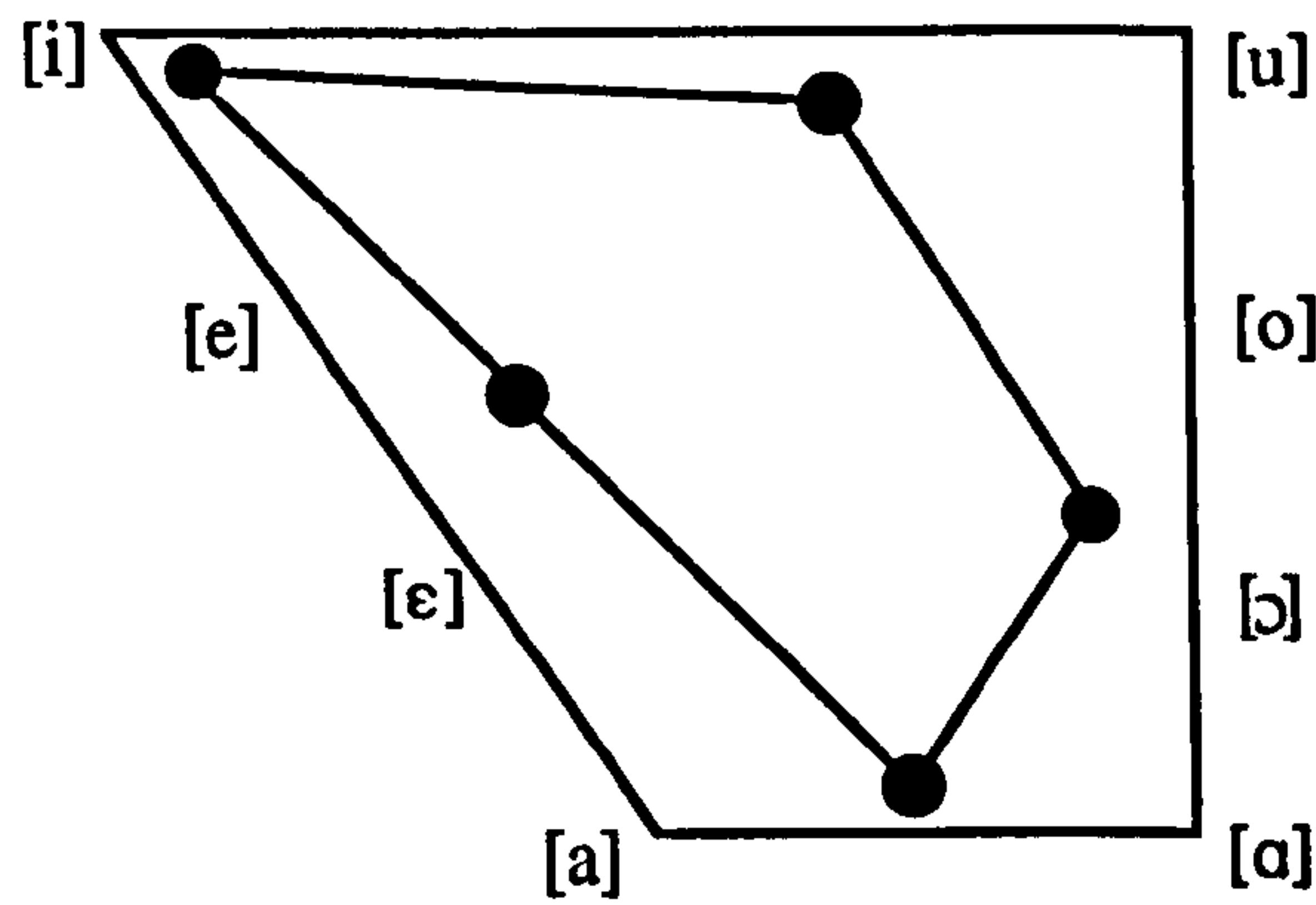
Among the vowels, only high short vowels /i/ and /u/ are commonly subject to the devoicing process, and vowel devoicing is a typical characteristic of north-eastern dialects including Standard Japanese (See Chapter 4, Section 4.4).

**Table 3.2    The Standard Japanese vowel system**

	- back	+ back	
+ high	i	u	- low
- high	e	o	
	a		+ low

The vowel positions are shown approximately in Figure 3.2

**Figure 3.2    Japanese vowel positions compared with the IPA cardinal vowels (based on Vance, 1987: 11)**



Standard Japanese /i/ has almost the same quality as IPA [i], but is slightly more open and less spread. Standard Japanese /e/ falls between IPA [e] and [ɛ]. The quality of /a/ falls roughly between [a] and [ɑ]. /o/ falls between [o] and [ɔ], but has very little lip rounding. For /u/, IPA [ɯ+] is the closest symbol.

The high vowels /i/ and /u/ are subject to vowel devoicing under normal circumstances, either between voiceless consonants or between a voiceless consonant and a pause. This can occur both within a word and across a word boundary, but devoicing does not occur between a pause and a voiceless consonant, e.g. /#usi#/ [ɯɕi]<sup>5</sup> 'cattle', /#iseki#/ [iseki] 'ruins'. The main reason why the high vowels are subject to vowel devoicing in Japanese as in other languages (see Chapters 2 and Section 5.4.1.1) is the intrinsically short duration of the vowels. Details of other conditions for vowel devoicing will be discussed in Chapter 4.

### 3.5 Consonant

#### 3.5.1 Phonemes and allophones

Modern Standard Japanese has sixteen consonantal phonemes as listed in (10). They are categorised into three groups: (1) consonants, (2) semivowels and (3) moraic consonants. Moraic consonants will be discussed in later in this section. As mentioned earlier in the previous section (Section 3.4), vowel devoicing occurs when a high vowel is either surrounded by voiceless consonants or preceded by a voiceless consonant and followed by a pause.

- |                      |                                      |
|----------------------|--------------------------------------|
| (10) a. consonants   | /p, b, m, n, d, t, z, s, r, h, g, k/ |
| b. semi-vowels       | /j, w/                               |
| c. moraic consonants | /N, Q <sup>6</sup> /                 |

In terms of /V-C-V/ coarticulation, in Japanese, the influence of a following vowel on its preceding consonant is generally stronger than that of a preceding vowel on its following consonant (Hirose, 1989: 82). As a result, Japanese consonants have many allophones, the exact nature of which are dependent on the following vowel. Therefore, when vowel devoicing occurs, it is still easy to identify the mora with a devoiced vowel because of the allophonic variation of the preceding consonant.

---

<sup>5</sup> The symbol [.] underneath a vowel means the vowel is voiceless.

<sup>6</sup> The symbol /Q/ denotes a moraic obstruent, namely the first element of geminate consonants.

Especially, the obstruents /s, z, t, d, h/ are subject to palatalisation and affrication, depending on the following vowels as shown in Table 3.3.

**Table 3.3 Allophones of the alveolar consonants in Japanese**  
(Originally from Shibatani except /h/, 1987: 137-138, but slightly modified)<sup>7</sup>

Consonant phonemes	Following vowels		
	before /i/	before /u/	elsewhere
/s/	[ɕ]	[s]	[s]
/z/	[dz] <sup>8</sup>	[dz] <sup>9</sup>	[z] <sup>10</sup>
/t/	[tɕ]	[ts]	[t]
/d/	[dz] <sup>2</sup>	[dz] <sup>3</sup>	[d]
/h/ <sup>11</sup>	[ɕ]	[ɸ]	[h]

This allophonic alternation is compulsory, and also occurs across a morpheme boundary (see Appendix I, Table 1). The same rule is also applied to foreign loan words, e.g. /siisoo/ [ɕi:so:] ‘seesaw’, /tippu/ [tɕip:ɯ] ‘tip’, /tuiido/ [tsui:do] ‘tweed’.

Recently borrowed foreign words have created greater numbers of possible phoneme sequences, and have been filling the phoneme inventory in Japanese. Conservative Japanese consonant and vowel sequences are listed in Appendix I, Table 2, and the examples of innovative CV sequences are listed in Appendix I, Table 3.

<sup>7</sup> Shibatani transcribed the allophones before /i/ not as alveolo-palatal but as palato-alveolar sounds: /s/ before /i/ as [ʃ], /z/ and /d/ before /i/ as [dʒ], /t/ before /i/ as [tʃ], which are not quite accurate descriptions, because these allophones lack lip-rounding. Therefore, I modified these descriptions to alveolo-palatal fricative and affricates.

<sup>8</sup> /di/ and /zi/ were historically distinct, but are neutralised as [dzi] in modern Standard Japanese.

<sup>9</sup> /du/ and /zu/ were likewise distinct in the past, but now are neutralised as [dzɯ].

<sup>10</sup> Some linguists consider the most common realisation of /z/ before /e, a, o, u/ is affricate /dz/ rather than fricative /z/.

<sup>11</sup> /h/ may optionally be voiced when preceded by a vowel and followed by a non-high vowel.

[h] ---> [ɦ] / V      V  
                                  [-high]



### 3.5.2 Moraic nasals

As briefly mentioned in Sections 3.2 and 3.5.1, the nasal consonants /m, n, ŋ/ consist of a particular category called the 'moraic nasal' when they occur syllable finally. This is because they consist of a mora on their own, and their place of articulation is dependent on the following sound: when it is followed by a consonant, it takes the same place of articulation as the following consonant, and when it is in utterance final position or in utterance internal position and followed by a vowel, it is realised as unreleased uvular nasal [ŋ]<sup>12</sup>. The moraic nasal in various phonetic environments are listed in Appendix I, Table 4.

Many studies (Sato, 1993, Han, 1994, etc.) reported longer duration of moraic nasals than non-moraic nasals in CV morae. However, durations of moraic nasals are not as long as the duration of CV morae. Durational aspects of moraic nasals and morae will be discussed in Chapter 7.

### 3.5.3 Moraic obstruents

The Japanese syllable structure does not normally allow consonant clusters. However, as briefly mentioned in Sections 3.2 and 3.5.1, voiceless obstruents<sup>13</sup> can consist of a mora on their own when they occur syllable finally as the first part of geminate consonants. Phonologically, geminate consonants may be analysed as sequences of two identical consonants. The first element of the geminates is moraic and often represented by the phonemic symbol /Q/. This is because, similarly to

---

<sup>12</sup> The IPA symbol uvular nasal with no audible release is [N]. However, as the symbol [ɳ] is used to mark an accent position in this thesis, I use an older version of IPA symbol for an unreleased uvular nasal [ŋ].

<sup>13</sup> In the modern innovative variety of pronunciations, voiced obstruents can be geminated in foreign loan words; for example, /be.d.-do/ [bed:o] 'bed', /ba.g.-gu/ [bag:u] 'bag', /dja.z.-zi/ [dzad:zi] 'judge'. However, voiced obstruent geminations are limited to certain loan words or onomatopoeia and have not established their status yet. Many Japanese speakers feel it more natural to substitute the voiceless counterpart of the obstruent in geminations, especially in the case of voiced plosive geminations, such as /be.t.-to/ [bet:o] for 'bed' and /ba.k.-k.u/ [ba.k:u] for 'bag'.

moraic nasals, the place of articulation of the moraic obstruents can be predicted from the place of articulation of the following consonant.

In fact, geminate consonants are phonetically long consonants as shown in (11), and an oral closure is not released until the release of the second element of geminates. Phonologically, as indicated above, they may be considered as sequences of the same voiceless consonants, because there is a syllable boundary between the geminate consonants, and they are separate morae (see Section 3.2).

(11) a.	/i.Q.-pa.i/	[ip:ai]	‘a lot’
b.	/i.Q.-ta.i/	[it:ai]	‘in what circumstance’
c.	/i.Q.-ti/	[it:ɕi]	‘accordance’
d.	/i.Q.-tu.i/	[it:sɯi]	‘a pair’
e.	/i.Q.-ka.i/	[ik:ai]	‘the first floor’
f.	/i.Q.-sa.i/	[is:ai]	‘everything’
g.	/i.Q.-sjo/	[iɕ:o]	‘together’

Moraic consonants occur only in syllable final, word internal position, and do not occur in other positions and their quality is totally predictable from the following consonants. Therefore, I will phonemicise moraic consonants as the same as the following consonant, e.g. /ja.t.-to/ ‘finally’, /ri.p.pa/ ‘magnificent’.

This brief explanation of basic Japanese phonology should be helpful in following discussion on Japanese vowel devoicing in the later chapters.

## Chapter 4

# Review of past work on Japanese vowel devoicing

### 4.1 Introduction

As discussed in Chapter 2, vowel devoicing can be found in many languages: in fact it varies from language to language, and can include vowel reduction and vowel deletion, and it is often associated with speech style, especially fast and casual speech. Generally a weak unstressed vowel tends to be devoiced or deleted. The most common devoiced vowel is a mid-central vowel, but high front vowels also undergo the "devoicing" process. This is due to the intrinsic shorter duration of high vowels which is conditioned physiologically by the degree of opening. (For full discussion, see Chapter 2.3).

The history of vowel devoicing in Japanese is not very clear. The earliest mention of vowel devoicing seems to be in "*Ars Grammaticae Iaponicae Linguae*" by D. Collado (1632) (Miyajima, 1961: cited by Maekawa, 1989: 139). However other historical facts are difficult to find. Most linguists treat voiceless vowels as allophones of voiced vowels. Martin (1952) is possibly the only person who has treated them as separate phonemes /i/, /j/, /ɯ/ and /ɯ̥/. The presence or absence of voicing of vowels does not alter the meaning of words. However, many linguists agree that in Standard Japanese vowel devoicing is not an optional fast speech rule nor in free variation with corresponding voiced vowels, but is rather a compulsory rule for the high vowels /i/ and /u/ in certain phonetic environments (Lovins, 1976; Sakurai, 1985; Vance, 1987;

Maekawa, 1989, etc.). A typical environment for vowel devoicing is when the high vowels are unaccented and preceded by a voiceless consonant and followed by a voiceless consonants or a pause.

(1)

$$\begin{array}{c} \text{V} \\ [+high] \\ [-accent] \end{array} \longrightarrow [-voice] \quad / \quad \begin{array}{c} \text{C} \\ [-voice] \end{array} \longrightarrow \left\{ \begin{array}{c} \text{C} \\ [-voice] \\ \# \end{array} \right\}$$

This rule is broadly correct, but rather too simplified. Details of the devoicing environment will be discussed later in this chapter.

Vowel devoicing phenomena occur not only in Standard Japanese, but also in many dialects, though it is particularly common in the eastern part of Japan (see Figure 4.1 in Section 4.5.9). Many studies have demonstrated that voiceless vowels can be found in many dialects which are not commonly thought to favour vowel devoicing such as the Kansai dialects and Tottori dialect (Sugito, 1987, and Maekawa, 1989).

## 4.2 Phonological and phonetic definitions of voiceless vowels

Maekawa (1989: 136) defines "vowel devoicing" as follows: "phonologically speaking, it is the phenomenon where there is no vocal fold vibration where there is supposed to be a vowel". His definition of a "voiceless vowel" is that it is "a vowel which is produced without vocal fold vibration, but has the same resonance and communicative features as the corresponding ordinary *voiced* vowel; i.e. the position of the tongue in the oral cavity is the same as the corresponding *voiced* vowel" (Maekawa, 1989: 138 [my translation]). Catford (1977: 120) defined voiceless close vowels as "voiceless approximants" describing them as sounds which maintain the same tongue-articulation as a voiced vowel, but without vocal fold vibration, giving a turbulent noise similar to voiceless fricatives.



Different analyses are needed for vowel devoicing depending on whether a vowel is preceded by a fricative or a plosive. Kawakami defined voiceless vowels as follows: "as long as conditions such as accent permit, morae /ki, pi/ and /ku, pu/ do not have a vowel when followed by a voiceless consonant. Instead, they have voiceless vowels [j] and [ɯ̥]. Morae /si, ti, hi/ ([ç̥i, t̥çi, ç̥i]) and /su, sju, tu, hu/ ([sɯ̥, çɯ̥, tsɯ̥, φɯ̥]) when followed by a voiceless consonant in general do not even have a voiceless vowel. Even if there *is* a vowel, its duration is extremely short" (Kawakami, 1977: 26 [my translation]. A fricative followed by a vowel, and a plosive followed by a vowel show different transitions from the consonant to the vowel. Plosives have three stages of articulation: (i) approach, (ii) closure and (iii) release. In articulatory terms, when a voiced plosive is followed by a vowel, the plosive is released into the following vowel, i.e. the vowel starts immediately after the release of the plosive. When a voiceless plosive is followed by a vowel, the plosive is also released into the following vowel, but the onset time and formant frequencies of the vowel is dependent on the degree of aspiration of the preceding plosive. When a vowel following a voiceless plosive is devoiced, there is no direct evidence of the following vowel, but the aspiration-like period after the release of the plosive has its resonance features of the voiceless counterpart of the voiced vowel.

When a vowel following a voiceless fricative is devoiced, however, it is not easy to draw a boundary between the fricative and the voiceless vowel. Fricatives are continuants produced with a narrow stricture. The tongue position for high vowels is very close to the roof of the mouth, creating a narrower passage for air to escape than for non-high vowels, but slightly more open than for fricative consonants. When a fricative is followed by a high vowel, their articulatory positions are often similar. The high vowels /i/ and /u/ are phonetically very similar to the consonants [j] and [ɣ]: producing [j], if we open the articulation, we will get [i], and also from [ɣ] we will get [ɯ̥]. In the same way, we can get [j] from [ç̥] and [ɯ̥] from [x]. Acoustically a fricative has an aperiodic waveform, whereas a voiced vowel has a periodic waveform. However, a voiceless high vowel has an aperiodic waveform which is similar to the turbulent noise of fricatives. Therefore, it is very difficult to separate a preceding voiceless fricative and a following voiceless vowel as they integrate into

each other. In most Japanese dialects, the influence of a following high vowel on the place of articulation of a preceding consonant is strong. As explained in Chapter 3, Table 3.3<sup>1</sup>, consonants are strongly palatalised when followed by a high front vowel /i/, and plosives are usually affricated when followed by the high vowels /i/ and /u/. An extreme case is /h/: when followed by /i/, glottal [h] becomes palatal [ç], and when followed by /u/, glottal becomes bilabial [ɸ].

When a vowel is devoiced, the articulatory change of a preceding consonant still occurs. Therefore it is still possible to identify an underlying vowel. /s/ in /si/ with a voiceless /i/ ([ç̥i]) is easily distinguishable from /s/ in /su/ with a voiceless /u/ ([sɯ̥]) because of the palatal articulation of the fricative, similarly /h/ in /hi/ with a voiceless /i/ ([ç̥i]) involves a positive movement of the tongue blade towards the hard palate, and also /h/ in /hu/ with a voiceless /u/ ([ɸɯ̥]) is produced with a positive lip closing action rather than a glottal consonant [h].

One problem of the coarticulatory effect on a preceding voiceless consonant of a following voiceless vowel is the case of [ç]. [ç] occurs not only as an allophone of /s/ before /i/ as explained above but also before /u/ as a realization of /sj/ in /sju/ sequences, e.g. /si̥'ki/ [ç̥i̥'ki] 'four seasons' vs. /sju̥'ki/ [çɯ̥'ki] 'memorandum'. In this case, too, there is coarticulation of a consonant and a following vowel and the difference in the articulation of the consonant is obvious; [ç] in [çi] is articulated more front than [ç] in [çɯ].

Beckman and Shoji (1984) carried out a perception experiment of consonant [ç] followed by voiceless high vowels [i] ([ç̥i]) and [ɯ] ([çɯ̥]), and reported that native speakers of Japanese could identify the underlying vowel following the fricative at better than chance level because of the spectral colouring of the devoiced vowels to the preceding fricative [ç], although spectrograms of the sequences showed only the frication noise of the fricative and did not show any formant-like characteristics of the following devoiced vowels. Tsuchida (1994) measured spectral peak frequencies of

---

<sup>1</sup> Other consonants are also palatalized before /i/, although broad phonetic transcription does not differentiate the symbols, e.g. [çi] [kɰ̥i], [ɲi] [ɲɰ̥i], etc.

the fricatives in /si/ and /sju/ sequences with voiced and devoiced vowels, and reported that although the fricative did not always have a spectral peak when the following vowels were devoiced, the average spectral peak frequencies of the fricatives were: 2490 Hz for [ç(i)], 2410 Hz for [ç(j)], 2110 Hz for [ç(ɯ)] and 2190 Hz for [ç(ɯ̥)]. The peak frequency was significantly higher for the fricatives followed by the vowel [i] regardless of voicing status of the vowel, but voicing did not have any effect on the peak frequency. Tsuchida concluded that when a devoiced syllable had a spectral peak, its frequency patterned in the same way as in voiced syllables, and thus there is a clear fricative-vowel coarticulation effect in Japanese even when the vowel is devoiced.

### 4.3 The physiological mechanism of voiceless vowel production

Sawashima (1969 and 1971) studied the state of the larynx during the production of voiceless vowels and not surprisingly found significant differences in the larynx for voiceless vowels compared with voiced vowels. Using photoelectric glottography, Sawashima (1969) found that there was no difference in the state of the glottis between a voiceless consonant and a voiceless vowel, although the glottal width for voiceless consonants varied considerably depending on the type of phoneme and their phonological environment.

The process of opening the glottis for the voiceless consonant of the first syllable is similar regardless of whether the following vowel is devoiced or not. Where the following vowel is devoiced the glottis remains wide open and seems to move immediately to a state appropriate to the following consonant (Sawashima, 1969: 35).

A further study by Sawashima (1971) on movement of the glottis with a fiberscope discovered that the opening of the glottis for a voiceless vowel was very similar to that for a voiceless consonant.



The glottal adjustments for devoicing of the vowel are not a mere skipping of the phonatory adjustments for the vowel but a positive effort of widening of the glottis for the devoiced vowel segment, even though there is no phonemic distinction between the voiced and devoiced vowels (Sawashima, 1971: 13).

However, the state of the glottis during production of voiceless vowels was different from the state of glottis during production of an intervocalic /h/. In Japanese, /h/ is usually voiced in an intervocalic position (Vance, 1987: 21)<sup>2</sup>. An initial /h/ in (2a) and (2b) is voiceless, but the second /h/ in (2a) is voiced because it is surrounded by the vowel /a/. When a polite marker /o/ is added in front of /hasi/ in (2b), the initial /h/ becomes intervocalic, thus it usually becomes voiced.

- |         |              |         |               |                                       |
|---------|--------------|---------|---------------|---------------------------------------|
| (2) (a) | <u>haha</u>  | /ha`ha/ | [ha`fia]      | 'mother'                              |
| (b)     | <u>hashi</u> | /ha`si/ | [ha`çi]       | 'chopsticks'                          |
|         |              | --->    | <u>ohashi</u> | /oha`çi/ [oña`çi] (o = polite marker) |

Yoshioka (1981) studied the the movements of the posterior cricoarytenoid muscle (PCA) which is the abducting muscle and the interarytenoid muscle (INT) which is the adducting muscle using laryngeal electromyography and fiberoptic endoscopy during the production of voiceless vowels and intervocalic voiced /h/. The test words used were /C1iC2ee/, /C1eeC2i/ and /C1eeC2ee/ where /C1/ and /C2/ were /h/ or /s/, i.e. /hihee/, /hisee/, /sihee/, /sisee/, /heesi/, /seesi/, /heehee/, /hesee/, /seehee/ and /sesee/ embedded in a carrier sentence (/i/s in bold italics are devoiceable, and the underlined /h/s are voiceable). As for /hihee/ and /hisee/, if the vowel /i/ remained voiced, the following /h/ was likely to become voiced, but if the /i/ became voiceless, the /h/ remained voiceless. The presence or absence of accent on vowels is also crucial in terms of voicing and devoicing of high vowels.

<sup>2</sup> When /h/ is followed by high vowels /i/ and /u/ such as /kahi/ 'for or against' and /sohu/ '(my) grandfather', the /h/ is not voiced even in an intervocalic position. Firstly, the change of articulation of /h/ from glottal [h] to [ç] before /i/ and to [ɸ] before /u/ occurs, then secondly devoicing of a following vowel occurs if a voiceless consonant or a pause follows after the vowel, e.g. /kahuku/ [kaɸɯ.kɯ]. 'disaster and happiness'. If neither of these processes is applicable, the /h/ may become voiced intervocalically.



Both the PCA and the INT showed different patterns of movements for voiced high vowels from devoiced high vowels. During the production of a devoiced vowel, the activation of the PCA with the suppression of the INT was observed; PCA activity increased rapidly for the initial voiceless consonant and continued to increase for the following devoiced vowel, and the INT activity curve showed one large continuous dip around the onset of the voiceless vowel. On the other hand, the results did not show any difference in the activities of both the PCA and the INT between voiced /h/ and voiceless consonants including /h/; there was clear opening movement of the PCA and no positive closing action of the INT even when vocal fold vibrations were observed during the production of an intervocalic /h/. Yoshioka suggested that the voicing of /h/ may have resulted from inevitable physiological processes of the glottis caused by glottal coarticulation between /h/ and surrounding vowels and had little to do with the speaker's neural control. On the other hand, the devoicing of vowels may be intentionally controlled by the speaker.

Voiceless vowels are differentiated from whispered vowels because they show different laryngeal gestures. Weitzman et al. (1976) investigated the vowel devoicing process in normal speech and two types of whispered speech, namely soft whisper mode and stage whisper mode<sup>3</sup> in Japanese, and concluded that the production of a voiceless vowel was characterised by the abduction of the vocal folds whereas a whispered vowel had a constricted larynx. Therefore it is possible to "devoice" vowels in whispered speech.

In normal speech, the take-off point of devoicing is from an adducted, voicing state, while in both types of whispered speech the take-off point is from an open but highly narrowed glottal state and a highly constricted supraglottal laryngeal state. .... Laryngeal constriction is a basic component in whispered speech and is maintained to a certain degree even during the devoicing gesture (Weitzman et al. 1976: 65).

Acoustic characteristics of voiceless vowels will be discussed in Chapter 5.

---

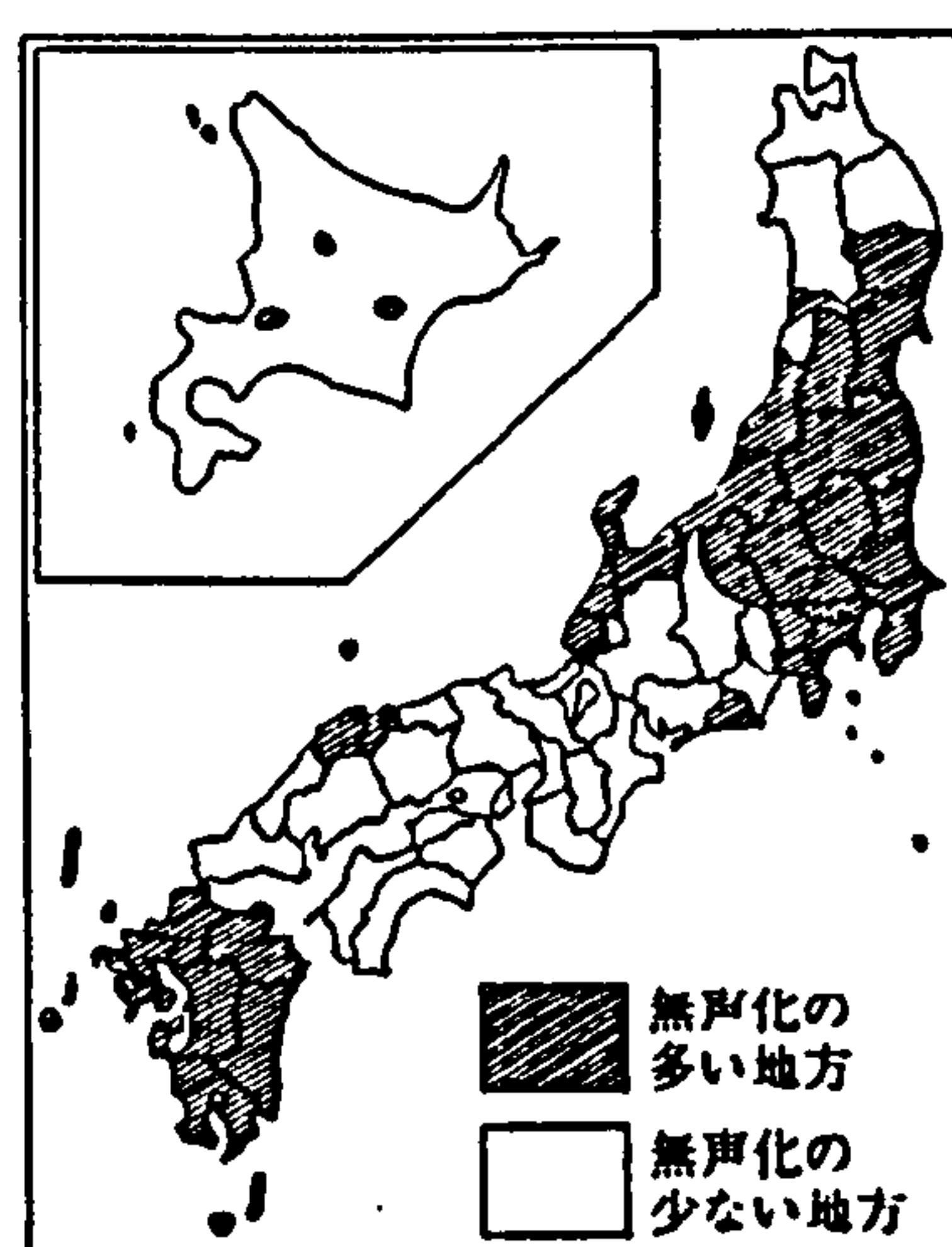
<sup>3</sup> "soft whisper mode" is defined as "the kind of whisper one ordinarily uses when whispering into someone's ear", and "stage whisper mode" is defined as "the kind of whisper one would use if the listener was quite a distance away from the speaker." (Weitzman et al, 1976: 62)

#### 4.4 The sociolinguistic function of vowel devoicing

As shown in Figure 4.1, vowel devoicing occurs mainly in north-eastern dialects and the Kyuushuu dialects. However, many devoicing cases have been reported in many other dialects where vowel devoicing is less likely to occur. For instance, Maekawa (1989) carried out research in the Tottori area where vowel devoicing is usually thought to be rare, and found that devoicing was in fact fairly common in all generations. Sugito (1988) also reports a high devoicing rate in the Nagoya dialect in which vowel devoicing is also supposed to be rare.

**Figure 4.1 Distribution of voiceless vowels**

- ◐ areas where vowel devoicing is common
- areas where vowel devoicing is not common



(From The NHK Accent Dictionary, 1985: 68)

Maekawa (1989: 143) pointed out that although the voicing of a vowel does not have a contrastive function there is a sociolinguistic implication in vowel devoicing, but not in voicing of /h/. For instance, it may sound strange for a speaker of Standard Japanese to pronounce the last mora with a voiced /u/ in "*hai, soudesu*" [ha'i so':desu] 'yes, it is' with a normal falling intonation. In Standard Japanese, when a vowel in the devoicing environment remains voiced, it sounds strange, or is an indication of another dialect or foreign accent. He also suggested that in the case of potential environments for consecutive vowel devoicing in a word such as *shukufuku* /sjukufuku/ (no accent) 'blessing' where all the /u/ vowels can potentially be devoiced,

although theoretically there were 16 possible voicing/devoicing patterns, in fact [ɕʌkʌɸʌkʌ], [ɕʌkʌɸʌkʌ] were possibly common but [ɕʌkʌɸʌkʌ] would sound odd. Consecutive devoicing cases will be discussed in detail in Section 4.5.8 and the later chapters.

Maekawa (1989: 143) also pointed out that vowel devoicing does not affect the meaning of words, but often provides other sociolinguistic information concerning, for instance, the area a speaker comes from, age, sex, mental state at the moment of speech (nervous, relaxed, etc). Thus, vowel devoicing does not seem to be a mere simplification of the phonation, but instead it is an intentional act of the speaker which carries sociolinguistical implications.

Regarding the generation difference, Mineta's study of the Shitara dialect in Aichi Prefecture showed a devoicing rate in the CVC environment of 82.3% in the younger generation and 61.9% in the older generation. The low percentage in the older generation was due to the extremely low devoicing rate (30%) in the accent nucleus position, which was about half of the rate for the younger generation (Mineta, 1988 cited by Maekawa, 1989: 151).

High vowel devoicing may be primarily caused by phonetic factors, i.e. glottal assimilation between inherently short high vowels and adjacent voiceless consonants (for details, see Section 4.5.1). As discussed earlier in Section 4.2, it is also greatly helped by the apparent articulatory change of a preceding consonant due to coarticulation between the consonant and the following vowel, which enables us to perceive the sound sequences even after the vowel is devoiced. An interesting point is that these physiologically conditioned processes occur in some dialects but not in others. Dialects in which vowel devoicing is common and dialects in which vowel devoicing is not common have the same sound system regarding the consonants and vowels in question. If vowel devoicing is purely a result of glottal coarticulation, it should be a common phenomenon in all dialects. Why devoicing is common in the Eastern dialects and less common in the Kansai dialect needs to be examined. Perhaps



devoicing is a more phonologically constrained process rather than a purely physiologically conditioned process.

#### 4.5 Factors affecting vowel devoicing in Japanese

As mentioned earlier, Japanese vowel devoicing involves the high vowels /i/ and /u/, which are usually devoiced when they are either between voiceless consonants or between a voiceless consonant and a pause. However, this process can be more complicated when several phonological factors interact. The effect of each factor may vary, and the relationship between factors and the hierarchical influence of the factors have not been well investigated.

In the NHK Accent Dictionary (1985: 128-129), Sakurai summarised vowel devoicing in Standard Japanese and suggested two rules and two tendencies.

- (3) **Rule I** : Vowels /i/ and /u/ are devoiced when they are surrounded by /p, t, k, s, h/ or preceded by one of these and followed by a pause.
- Rule II** : Unaccented vowels /i/ and /u/ in the first mora of an utterance are devoiced when they are preceded by a pause.
- Tendency I** : Unaccented /a, o/ in morae /ka/ and /ko/ are likely to be devoiced when an identical mora follows.
- Tendency II**: Unaccented /a, o/ in morae /ha/ and /ho/ are likely to be devoiced when the vowel in the following syllable is the same vowel.

However, he added that in actual conversation vowel devoicing does not always abide by these two rules and tendencies, and that there were a lot of exceptions.

The main factors which have been reported to have an effect on vowel devoicing in Japanese are: (a) the type of vowels, (b) the preceding consonant, (c) the following consonant, (d) the vowel in the following mora, (e) pitch and accent, (f) position in the word, (g) breath group, (h) speech style, (i) speech rate, (j) regional and



generation differences, and (k) grammatical structure of the word. A number of studies have been carried out to investigate vowel devoicing in Japanese, and the effects of some of these factors have been well documented, but little research has been carried out concerning the effect of other factors.

Most studies have focused on the effect of one particular factor on devoicing. To evaluate the effect of these factors, all studies employed devoicing rates on each factor, i.e. the percentage of devoiced vowels for each factor. Spectrographic analysis was usually used to obtain devoicing rates based on the presence or absence of voice bars on a wide band spectrogram, or on the presence of fundamental frequency and harmonics on a narrow band spectrogram.

The data on which these studies were based ranged from one subject to 29 subjects, and from a few words to a large scale Japanese speech database which was composed of phonetically balanced words and sentences. However, in all the studies a categorical distinction was made between voiced and voiceless vowels. The development of facilities such as spectrograph has made the analysis of vowel devoicing easier and clearer. However, it has also introduced a new problem when analysing devoicing, in that spectrograms sometimes show a steady gradient in devoicing (see Chapter 5, Section 5.4.1). However, all previous studies categorised vowels into only two groups, namely voiced and voiceless. None of the studies I know of have discussed intermediate cases. Each study must have had criteria for determining the devoicing of vowels, but these criteria were never mentioned. My data, described in the later experimental chapters, show that although there were variations in the degree of devoicing, most vowels fell clearly into either voiced or voiceless categories. I assume that most data in the studies quoted below could be fairly clearly distinguished as voiced or voiceless, and if not, the researchers must have categorised them according to their own criteria, which may have varied between studies.

In this section, I will discuss in detail the effect of the above factors and some problems with regard to vowel devoicing.

## 4.5.1 Phonetic environment

### 4.5.1.1 The vowel itself

Past studies have shown that both high and non-high vowels can be devoiced, but factors which condition devoicing for these two types of vowels may be different. As mentioned in Chapter 3, Section 3.4.1, vowel devoicing mainly involves the high vowels /i/ and /u/. This may be due to the inherent shorter duration of high vowels. Han (1962b) showed big durational differences for Japanese vowels: in her work the average durations of the Japanese vowel /u/ (which is the shortest) was taken as 1.00 as a unit of reference, and the average duration of the other vowels were /i/ 1.17, /o/ 1.26, /e/ 1.37 and /a/ 1.44. Vance (1987: 49) quoted Sakuma (1922) stating that /i/ and /u/ were more susceptible to devoicing than /a, e, o/ because they were intrinsically shorter: it seems to be language universal that everything being equal, the lower a vowel, the longer its intrinsic duration.

As for the difference of devoicing between /i/ and /u/, Han (1962a: 84-85) claimed that /u/ was more likely to be devoiced than /i/, because /u/ is inherently shorter than /i/. She confirmed this statement in her experiments, but unfortunately her study was not very extensive; her results were based on one speaker's pronunciation of two pairs of words such as /kisi/ 'shore' - /kusi/ 'comb' and /pi'kupiku/ - /pu'kupuku/ 'both words are onomatopoeias'. Fukai (1979: 50) discussed vowel devoicing frequency in terms of vowel sonority. The sonority of a sound is its loudness relative to that of other sounds with the same length, stress, and pitch (Ladefoged, 1982: 221). Fukai suggested that back vowels were devoiced more often than front vowels because back vowels require less energy during their production due to the fact that the sonority of back vowels was less than that of front vowels. However, not all studies supported this view. For example, Maekawa (1989: 144) reported there was no difference in devoicing between /i/ and /u/. One of the difficulties in explaining vowel devoicing in terms of sonority is that generally sonorities of open vowels /a/ and /o/ are higher than that of /e/, but devoicing of /a, o/ has been reported to be more common than that of /e/.

Most studies agree that the non-high vowels /e, a, o/ can also be devoiced when they occur in the initial mora followed by an identical syllable. For instance, in the following words non-high vowels in the initial mora can be devoiced: kakaku [kəkakɯ] 'price', Sasaki [səsaki] 'a surname', sosogu [sɔsɔŋɯ] 'to pour', kokoro [kɔko'ro] 'heart', etc. /e/ seems least likely to be devoiced, but there are some examples of [ɛ]: kesshite [kɛɕ:ite] 'never', sekkaku [sɛk:akɯ] 'especially, kindly' (Sakuma, 1922 cited by Vance, 1987: 48, and Maekawa, 1989: 136). Other examples often mentioned are katana [kətana] 'sword', haka [həkə] 'grave', hokori [hɔkɔri] 'dust', sotobori [sɔtɔbɔri] 'outer moat', where the devoiceable vowels are followed by a mora consisting of a voiceless consonant and an identical vowel, but are not necessarily followed by the identical syllable.

However, the devoicing of non-high vowels is far less common than that of high vowels, and it seems to be sensitive to speech style and speaking rate in addition to ideal phonological and phonetic environments, i.e. adjacent to voiceless sounds. Yoshida and Sagisaka (1990) used the ATR Large Scale Speech Database<sup>4</sup>, and analysed 5449 isolated words and 75 read sentences by three male speakers, and reported only four examples of devoiced /a/ in utterances of isolated words by only one of their subjects. Bloch (1950: 136) listed examples of non-high vowel devoicing exclusive of /e/. Bloch differentiated high vowel devoicing and non-high vowel devoicing and said that high vowel devoicing could occur even in slow or careful speech, but that non-high vowel devoicing ([ɤ, ɔ]) could only occur in rapid speech. Han (1962a: 84-85) also stated that /e, a, o/ were not normally devoiced at a normal tempo. Martin stated that there was individual speaker variation in non-high vowel devoicing and that it usually occurred in an initial syllable followed by an identical syllable, as in /kokoro/ (Martin, 1952 cited by Vance, 1987: 49). Kawakami (1977: 69) also said that the devoicing of non-high vowels was optional and varied from

---

<sup>4</sup> ATR Large Scale Speech Database consists of approximately 5,000 common Japanese words, 200 phonetically balanced words, 100 Japanese CV syllables and small number of short sentences. The speech data were manually labelled with a narrow phonetic transcription by inspecting spectrograms.



speaker to speaker, whereas the devoicing of high vowels was obligatory except in unusually precise speech.

One case of supposedly obligatory non-high vowel devoicing is stated by Nihon Onsei Gakkai (The Phonetic Society of Japan) (1976), namely the sentence "*arimashita*" /arima'sita/ 'It was there', where /a/ in the last mora is usually pronounced without voicing [arima'çit̚] (Vance, 1987: 49). As far as I am aware, apart from the last example, the devoicing of non-high vowels occurs only when non-high vowels are surrounded by voiceless consonants. Maekawa (1989: 137) suggested that the devoicing of /e, a, o/ was not frequent and did not have any sociolinguistic significance. Therefore it should be treated as a different phenomenon from high vowel devoicing. Vance (1987: 49) said that Sakuma was the only linguist who claimed the devoicing of the first /o/ in /kokoro/ to be obligatory, while others treated non-high vowel devoicing as an optional casual and fast speech rule.

All the examples of words with non-high vowel devoicing cited above do not have an accent on the first mora. Most of them are in non-accented words. In other words, the first mora which is likely to be devoiced has lower pitch than the second mora at least in citation, i.e. the vowel in the first mora has less energy than the vowels in other morae. Similarly, the last vowel /a/ in the last example /arima'sita/ is in an unaccented position and the last mora has the lowest pitch in the phrase because of the downtrend (see Chapter 3 for the explanation of 'downtrend'). Therefore, it has less energy than other vowels. This may be one of the crucial factors for non-high vowel devoicing since non-high vowel devoicing seems to be restricted to the above environment. The relationship between pitch and vowel devoicing will be discussed in detail in Section 4.5.2.

#### 4.5.1.2 The vowel in a following mora

Maekawa (1989: 145) pointed out that the vowel in a following mora has some influence on vowel devoicing. He quoted the study of the Toohoku dialect by Inoue



(1968) and the Aichi dialect by Mineta (1988), and summarised that a high vowel was more likely to be devoiced when the vowel in the following mora was /a/, /e/ or /o/ rather than /i/ or /u/. Sugito (1988) also mentioned that in the Kansai dialect (in which vowel devoicing is less common) /u/ in /kusa/ (no accent) 'grass' is very often devoiced. Maekawa's later studies (1989 and 1990) also supported this claim.

Maekawa's study (1989) showed that the devoicing rate for /si/ was higher when the vowel of the following mora was [a], [e] or [o] than [i] in all tested environments of /\_ÇVC/, /\_ÇVC/ and /\_ÇVN/. In another study, Maekawa (1990) found reduction and devoicing of /u/ in the /ÇuÇV/ environment at different speaking rates. His results showed that both reduction and devoicing rates were higher when the vowel of the following mora was /a/ rather than /u/. Miyajima (1961) claimed that this influence of the non-high vowels in the following mora already existed in 17th century Japanese (cited by Maekawa, 1989: 146). The reason for this may be in order to avoid consecutive devoicing. However, this area has not been well studied.

#### 4.5.1.3 The preceding consonant

Consonants which precede voiceless vowels are predominantly voiceless obstruents. However some devoicing examples preceded by voiced consonants are reported. Yoshida and Sagisaka (1990) reported that one of their 3 subjects devoiced a vowel after [j] (2 examples out of 538 ) and another subject devoiced 12 examples out of 734 after [dz], [d] and [m].

Most studies agree that among voiceless consonants the vowel devoicing rate is higher when vowels are preceded by fricatives, as in /hi/, /hu/, /si/. For example, Takeda and Kuwabara (1987) used isolated words spoken by a male speaker in the ATR Large Scale Speech Database, and studied the devoicing rate for /i/ and /u/ in 3188 samples from 5626 isolated words and analysed the results by preceding consonant. Their results showed that the highest devoicing rate was in the syllable [çi] (56/138 samples [41%]), then followed by [ɸɯ] (45/156 samples [29%]) and [çi] (116/499 samples [23%]). In terms of the sample size, [çi] and [tsɯ] occurred most often (116/499

samples and 90/499 samples respectively). Their results will be discussed in detail in Chapter 5.

Han (1962a) suggested a hierarchy of devoicing frequency by preceding consonants: "affricate and fricative > plosive"<sup>5</sup>. She concluded that the vowels preceded by non-stops were more readily devoiced than vowels preceded by stops (1962a: 91). The results of Yoshida and Sagisaka (1990: 1) showed a hierarchy of "affricate > fricative > plosive" in one of their subjects, but no specific tendency was found in the other two subjects. However, they report a different hierarchy for reading style: "fricative (65-68%) > plosive and affricate (24-27%)" (Yoshida and Sagisaka, 1990: 3).

A possible reason for this disagreement is the influence of other factors such as accent position, the following consonant, a combination of preceding and following consonants, position in a word, speech style, etc. These data were not sampled under controlled conditions, nor were these studies aimed at clarifying the effect of consonant types under various phonetic conditions. They simply categorised the devoicing rates by consonant types.

One certain vowel devoicing constraint is that devoicing never occurs when a vowel is preceded by a pause: " # \_\_\_\_ "

#### 4.5.1.4 The following consonant

The consonants following devoiced vowels are also predominantly voiceless obstruents. Many studies also find that vowel devoicing occurs when a following consonant is voiced. Yoshida and Sagisaka (1990) reported 62 samples of devoiced vowels followed by a voiced consonant out of 1873 CV-syllables. The most common voiced consonant was /m/ (32 samples), followed by /n/ (15 samples), /g/ (20 samples), /b/ (3 samples), /r/ /d/ (2 samples each) and /j/ /w/ (1 sample each).

---

<sup>5</sup> Here and thereafter, the symbol "A > B" means "A" is more likely to trigger devoicing than "B", or "A" is more likely to be devoiced than "B", etc.

However, they do not mention the combination of the types of preceding and following consonants. Takeda and Kuwabara (1987: 105) reported that the devoicing frequencies of /i/ preceded by /s/ were: 50% when followed by a voiced consonant: /g/ (5 examples), 15% when followed by /n/ (2 examples), and 3% when followed by /m/ (1 example).

Maekawa (1989: 144) concluded from two of his studies of /ki/ and /si/ morae (one study was based on one professional female announcer, and the other study was based on 29 male Tottori dialect speakers) that /i/ and /u/ could be devoiced when they were followed by a voiced consonant, although their devoicing rate was lower than cases followed by a voiceless consonant. His results showed that among voiced consonants, the vowel devoicing rate was highest for nasals (/m, n, ŋ/), followed by voiced plosives (/d, g/), a liquid (/r/), and semivowels (/j, w/). It must be noted that there was a high devoicing rate (about 40%) when [çi] was followed by /r/.

Fukai (1979: 47) stated that the pronunciation of furo /huro'/ (bath) was [φro'] in ordinary conversation, and that [φɯro'] was an example of extremely careful pronunciation.

There is a slight disagreement regarding the order of effect of following consonant types. Han (1962a) stated the order "affricate and fricative > plosive". Following their own research on [çi] in 1987, Kuwabara and Takeda (1988: 4) did a further study on the order of effect for [φɯ] using the ATR Large Scale Speech Database, and concluded that the order was "plosive > fricative > pause".

Yoshida and Sagisaka (1990: 2) showed similar results for isolated words: "plosive (86-99%) > affricate (73-93%) > fricative (51-60%)". No obvious individual speaker differences were found among their three subjects. In reading style, the order was similar: "plosive > fricative > pause (= utterance final)".

The crucial point here may be the combination of consonants. In the sequence of /C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>2</sub>/ where "V<sub>1</sub>" is potentially voiceable, it is important to take the



combination of C1 and C2 into account. With the exception of Han (1962a), none of the studies discussed the combinations of C1/C2, and even with Han's data, the influence of accent should have been taken into account. Takeda and Kuwabara (1987) and Kuwabara and Takeda (1988) both observed cases where C1 was a fricative ([ç] for Takeda and Kuwabara, [ɸ] for Kuwabara and Takeda). Yoshida and Sagisaka (1990) did not classify the data according to the C1 consonant types.

Most of the literature is in agreement that an utterance final pause functions in a similar way to voiceless consonants, and triggers vowel devoicing, but devoicing of the preceding vowel is less frequent than when the vowel is followed by a voiceless consonant (Kawakami, 1977, Sakurai, 1985, Vance, 1987, etc.). When a high vowel preceded by a voiceless consonant occurs in an utterance final and also a sentence final position, it is likely to be devoiced. However, the examples in the literature above seem to be more or less restricted to certain lexical items, such as -desu 'COPULA [POLITE]' and -masu 'POLITE'<sup>6</sup> both of which often appear sentence finally. When a devoiceable vowel is in utterance final and also phrase final position, devoicing of the vowel is rare. Maekawa (1989: 147) provided the following examples:

- (4) a. "Watashi ga jussai no tokii, (pause)"                      'When I was ten years old, ...'  
       b. "Zenzen ase mo kakanaishii, (pause)"                      '(I don't) sweat at all, and...'

The underlined /i/ in (4a) and (4b) are in final position in an accentual phrase, and are preceded by voiceless consonants ([k] in (4a) and [ç] in (4b)) and followed by a pause, but they are not normally devoiced. Phrase final non-devoicing may be closely related to the intonational pattern. In a declarative sentence in Japanese, pitch normally falls at the end of a sentence, whereas pitch either goes slightly up or stays at the level at the end of a phrase followed by a pause. Higher pitch may be a crucial factor for

---

<sup>6</sup> The definitions and functions of Japanese morphemes -desu and -masu vary enormously depending on scholars. Some syntacticians define both -desu and -masu as "auxiliaries", some define them as "modal verbs", others as "copulas", etc. Here I follow their definitions and translations by Shibatani (1990): -desu as 'COPULA [POLITE]' meaning "copula marking politeness", and -masu as 'POLITE' meaning "a morpheme marking politeness".



blocking vowel devoicing at a phrase and clause final position. The relationship between vowel devoicing and pitch will be discussed in detail in the next section.

Following geminate consonants, which are manifested as a simple continuation of turbulent noise for fricatives, or of silence for the stop closures (Sawashima, 1971: 8), do not seem to block vowel devoicing (Sawashima, 1969: 36, Lovins, 1976: 118, Fukai, 1979: 49, etc.). The NHK Accent Dictionary (1985) lists all devoiceable vowels before geminate consonants as voiceless.

#### 4. 5. 2 Pitch and accent

Most literature says that it is unlikely for a vowel to be devoiced when the vowel is accented<sup>7</sup>. Sakurai (1985: 129) noted the following exception to the vowel devoicing rule I (see Section 4.5), namely that "vowel devoicing does not occur when a high vowel is in a high pitched mora and the following mora is low pitched". He listed examples /i/ in /si'soN/ [ɕi'soŋ]<sup>8</sup> 'descendant' and the first /u/ in /kaku'su/ [kakɯ'sɯ] 'to hide' both of which are voiced in accented positions. He also noted exceptions to rule II (see Section 4.5) that "there are two possibilities of pronunciation for the case when a devoiceable vowel is accented: (a) an accented devoiceable vowel becomes voiceless, e.g. /itu'ku/ [itsɯ'kɯ] 'to remain staying', /garasu'ki/ [garasɯ'ki] 'glassware', and /sjoki'kaN/ [ɕoki'kaŋ] 'secretary', and (b) the accent shifts one mora rightwards and the vowel becomes voiceless, e.g. tsuku 'to arrive' /tu'ku/ ---> /tuku'/ [tsɯ'kuɾ], kite 'to come (interrogative)' /ki'te/ ---> /kite'/ [kɯ'te] and shiki 'four seasons' /si'ki/ ---> /siki'/ [ɕiki'].

Vance (1987: 50) studied Hirayama's pronunciation dictionary (1960) and concluded that there was a tendency to pronounce an accented mora with a voiced vowel, even in the devoicing environment, in order to avoid an accent on a devoiced mora. Han

<sup>7</sup> In Standard Japanese the word "accented" means there is a fall in pitch after a vowel, i.e. in the sequence of C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>2</sub>C<sub>3</sub>V<sub>3</sub> if there is a pitch drop between V<sub>2</sub> and C<sub>3</sub>, V<sub>2</sub> or the mora 'C<sub>2</sub>V<sub>2</sub>' is said to be accented. For further detail, see Chapter 2.

<sup>8</sup> /N/ and [ŋ] are phonological and phonetic symbols for moraic nasals.

(1962a) said that devoicing generally applied only to vowels in unaccented morae although there were some exceptions.

Takeda and Kuwabara (1987: 105) studied devoicing of the /si/ mora in various accent positions and showed that the vowel was most likely to be devoiced when a preceding mora was accented. Frequencies of devoicing were:

(5) a.	in an accented mora	0%	(0 case/121 samples)
b.	the preceding mora was accented	34%	(53 cases/155 samples)
c.	others	28%	(63 cases/223 samples)

Takeda and Kuwabara (1987) also analysed their data according to Sakurai's rule II (see Section 4.5) but their results differed from those of Sakurai. There was only one devoicing case out of 29 samples which fit the devoicing condition of a low pitched mora before a pause, while there were 115 devoiced cases out of 470 samples which do not fit this devoicing condition. They suggested that these differences might be attributed to different speaking styles: their results were based on isolated words, but Sakurai's results were based on continuous speech.

Kuwabara and Takeda (1988) showed that as far as devoicing in the mora /hu/ ([ɸɰ]) was concerned, devoicing was least likely when the mora was accented: their order of accent position is:

- (6) on the mora itself < on the preceding mora < on another mora.

Yoshida and Sagisaka (1990) obtained similar results. They studied vowel devoicing according to pitch<sup>9</sup>, accent in isolated words, and the number of morae in a word in reading style. All of their subjects showed more devoicing in a low pitched mora than in a high pitched mora, regardless of the number of morae in a word. In all cases, the devoicing of the vowel of a low pitched mora was more frequent than that of a high

---

<sup>9</sup> They used the terms "high accent" and "low accent" for this study. However I understand that they meant "high pitch" and "low pitch" respectively. Therefore I have used the term "pitch" here.

pitched mora except in one case where one subject showed the same devoicing rate (96%) for both high pitched and low pitched morae in five mora words. No individual differences were found. The devoicing rate for accented vowels was about 42% and this figure was lower than the lowest devoicing rate in high pitched morae (43% in the high pitched mora in the two mora words). It must be noted, however, that one of the subjects showed an extremely high devoicing rate (91%) for accented vowels when the following consonant was [k] but only 17% when followed by [s] or [ɕ]. They concluded that there must have been a strong influence of the type of following consonant on devoicing in the accented position. But clear individual differences were not evident with regard to devoicing in accented morae.

Lovins (1976) also studied devoicing rates in relation to pitch and accent. In her experiment, 10 Standard Japanese speakers (5 male and 5 female) pronounced 56 sentences representing 14 different utterance types of contrasting pitch patterns, e.g. /hasi to kaita/ '(He) wrote bridge', /hasi to kaita/ '(He) wrote edge', and /ha'si to kaita/ '(He) wrote chopsticks'. Her results were:

(7) a. in low pitched unaccented morae	100%,
b. in high pitched pre-accentual morae	more than 98%
c. in accented morae	76%

Lovins concluded that high or low pitch did not affect devoicing rate very much but accentuation did. Accented vowels were less likely to be devoiced. She argued that it was possibly because devoicing of a vowel caused partial deletion of the pitch contour and an accented mora without voicing might cause confusion. Since one cannot hear a pitch drop in a voiceless vowel, it would be unclear in a sequence like /#MMM̥M/ (where 'M' indicates a mora), or /LHHL/, if the third mora were perceived as pitchless, whether the accent really fell on the second or the third mora. This point will be discussed again in later in this section.

The results of the above studies suggest vowel devoicing does occur even in accented morae although less frequently than in low pitched or unaccented morae. Even the



NHK Accent Dictionary (1985) lists two possibilities of pronunciation, (a) with an accented devoiced vowel, and (b) an unaccented devoiced vowel for quite a few words, for example,

(8)	<u>word</u>	<u>gloss</u>	<u>(a)</u>	<u>(b)</u>
	<u>kisetsufuu</u>	'the monsoon'	[k̚isetsu̥ɸɯ:]	[k̚isetsu̥ɸɯ:] (no accent)
	<u>sashikizu</u>	'a stab'	[sa̠ɕi̥'kid͡zɯ]	[sa̠ɕikid͡zɯ] (no accent)
	<u>shuki</u>	'memorandum'	[ɕɯ̥'ki]	[ɕɯ̥ki] (no accent)
	<u>sappiku</u>	'to deduct'	[sap:i̥'kɯ]	[sap:i̥kɯ] (no accent)

Needless to say, it lists many possibilities for voiceless vowels in high pitched morae.

It has been said earlier that in order to avoid having an accent on a devoiced mora the accent would be shifted to another mora (Lovins, 1976: 118, McCawley, 1977: 266, Sakurai, 1985: 129, Maekawa, 1989: 146, etc). It has also been suggested that accent blocks devoicing. McCawley (1977) demonstrated the presence of the accent shift phenomenon in adjective past-tense formation. Japanese adjectives have the accent on the last mora of the stem in the infinitive form and on the penultimate mora of the stem in the past-tense form as shown in (9)<sup>10</sup>.

(9)	<u>adjective</u>	<u>gloss</u>	<u>infinitive</u>		<u>past-tense</u>
	/hiro-i/	'wide, large'	[ɕiɾo̥'-i]	----->	[ɕi'ɾo-kat-ta]
	/samu-i/	'cold'	[samɯ'-i]	----->	[sa'mɯ-kat-ta]
	/omosiro-i/	'interesting'	[omoɕiɾo̥'-i]	----->	[omoɕi'ɾo-kat-ta]
	/atataka-i/	'warm'	[atataka'-i]	----->	[atata'ka-kat-ta]

But in the following examples, the accent is not on the penultimate mora of the stem, but on the last mora of the stem as in the infinitive form as shown in (10).

<sup>10</sup> Morpheme boundaries are indicated with a symbol “-” in all examples of adjectives in this section. The morpheme /-i/ of the infinitive form of adjectives is inflectional. When a past tense marker /ta/ follows an adjective, the morpheme /-i/ changes to /-kat/.



(10)

<u>adjective</u>	<u>gloss</u>	<u>infinitive</u>		<u>past-tense</u>
/tika-i/	'near'	[tɕika'-i]	----->	[tɕika'-kat-ta]
/kasiko-i/	'clever'	[kaɕiko'-i]	----->	[kaɕiko'-kat-ta]
/huto-i/	'fat, wide'	[ɸuɽto'-i]	----->	[ɸuɽto'-kat-ta]

The accent in the past-tense form has apparently not shifted because the vowel in the penultimate mora of the stem is devoiced. In other words, the accent position in the past-tense form remains unchanged from the infinitive form because of the vowel devoicing of the preceding mora. Compare (10) with (11):

(11)

<u>adjective</u>	<u>gloss</u>	<u>infinitive</u>		<u>past-tense</u>
/atsu-i/	'hot'	[atsɯ'-i]	----->	[a'tsɯ-kat-ta]
/hosi-i/	'to want'	[hoɕi'-i]	----->	[ho'ɕi-kat-ta]
/atarasi-i/	'new'	[ataraɕi'-i]	----->	[atara'ɕi-kat-ta]

In these adjectives, the accent shifts to the penultimate mora of the stem and the vowel in the last mora is devoiced.

In the case of adjectives such as hikui /hiku'-i/ 'low' and utsukushii /utsukusi'-i/ 'beautiful', the NHK Accent Dictionary lists the pronunciation as [ɕikuɯ'-i] and [ɯtsuɕkuɯɕi'-i] respectively with no consecutive vowel devoicing. Their past-tense forms theoretically have two possibilities, i.e. [ɕi'kuɯ-kat-ta] or [ɕikuɯ'-kat-ta], and [ɯtsuɕkuɯ'ɕi-kat-ta] or [ɯtsuɕkuɯɕi'-kat-ta]. However in actual pronunciation there are other possibilities such as [ɕi'kuɯ-kat-ta], [ɕikuɯ'-kat-ta], [ɯtsuɕkuɯ'ɕi-kat-ta], [ɯtsuɕkuɯɕi'-kat-ta], etc. All examples in (10) and (11) above have alternative pronunciations without accent shift, i.e. with an accented voiceless vowel:

The accent shift phenomenon was more or less obligatory in the past, but it is said that the phenomenon has been weakened in recent years. Speakers do not necessarily shift the accent in order to avoid having the accent on a devoiced mora, but rather keep the

accent on the devoiced mora not only in the adjective past tense but also in all types of words as shown in (12).

(12)

<u>adjective</u>	<u>traditional past tense</u>	<u>alternative past-tense</u>
/tikakat-ta/	[tɕika'-kat-ta]	[tɕi'ka-kat-ta]
/kasikokat-ta/	[kaɕiko'-kat-ta]	[kaɕi'ko-kat-ta]
/hutokat-ta/	[ɸɯto'-kat-ta]	[ɸɯ'to-kat-ta]
/atu-katta/	[a'tɕɯ-kat-ta]	[atsɯ'-kat-ta]
/hosi-katta/	[ho'ɕi-kat-ta]	[hoɕi'-kat-ta]
/atarasi-katta/	[atara'ɕi-kat-ta]	[ataraɕi'-kat-ta]

N. Hattori (1989) carried out an experiment to show that the accent could fall on a voiceless vowel. She measured the fundamental frequency (F0) of two groups of words; (i) words in group 1 include a devoiceable vowel, and (ii) words in group 2 contain only voiced vowels. She compared the F0 contours of words in group 1 and 2, but the comparison was made only between words of the same number of morae between the two groups. For instance, she compared the F0 contours of 4 mora words such as /ho.o.ti .ki/ 'fire alarm', /bi.i.si .ki/ 'artistic sense'<sup>11</sup>, where /i/s in the third mora in both words (written in bold italics) are devoiceable and theoretically accented, with other sets of 4 mora words (a) /ha.N.bu'.N/ 'half', /ho.o.i'.do/ 'latitude' /ke.i.ri'.bu/ 'accountant's department', all of which have an accent on the third mora, and (b) /hi.ma'.wa.ri/ 'sunflower', /zi.mu'.i.N/ 'clerk', /no.mi'.mo.no/ 'beverage' which have an accent on the second mora. She also compared 5 and 6 mora words such as /zi.mu.ki .te.i/ 'business regulation' (5 morae), /i.rjo.o.ki .ka.i/ 'medical instruments' (6 morae), where also /i/ in the third mora in /zi.mu.ki .te.i/ and /i/ in the fourth mora in /i.rjo.o.ki .ka i/ are devoiceable and theoretically accented, with other sets of 5 and 6 mora words (a) (i) 5 mora words /ba.re.ri'.i.na/ 'ballerina', /ze.mi.na'.a.ru/ 'seminar' which have an accent on the third mora, (ii) 6 mora words /mo.no.no.a'.wa.re/ 'sentiments', /a.N.mo.ni'.u.mu/ 'ammonium' which have an accent on the fourth mora, and (b) (i) /bu.a'.i.so.o/ 'unsociable', /go.jo'.o.zi.N/

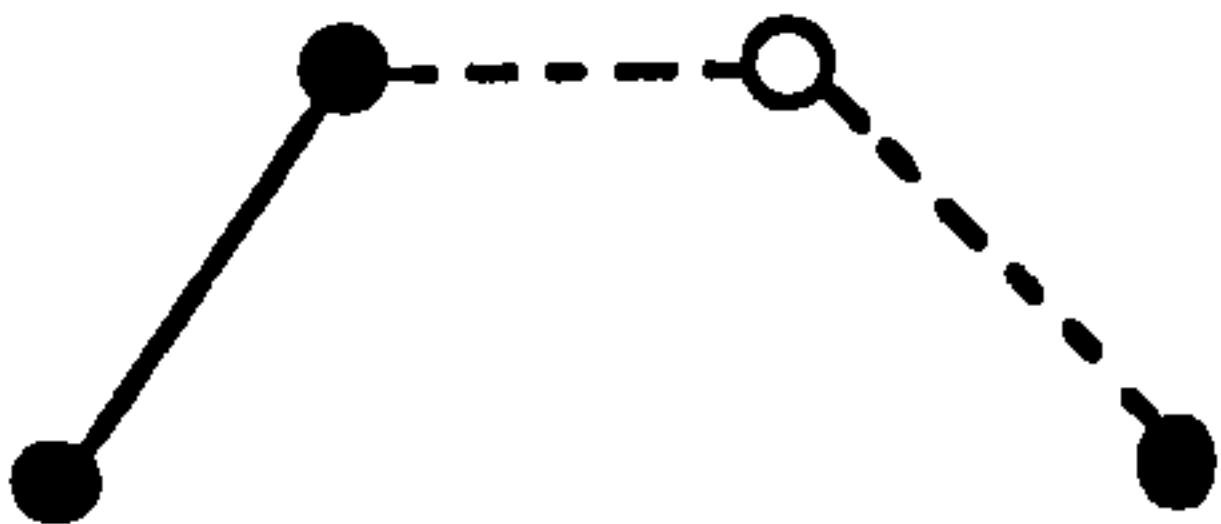
<sup>11</sup> Mora boundaries are indicated by a symbol /. /in this section.

'Beware of ---" which have an accent on the second mora, (ii) /sa.ra.ri'.i.ma.N/ 'office worker', /ju.u.me'.i.zi.N/ 'celebrity' which have an accent on the third mora as shown in Figures 4.2 (1) - (3).

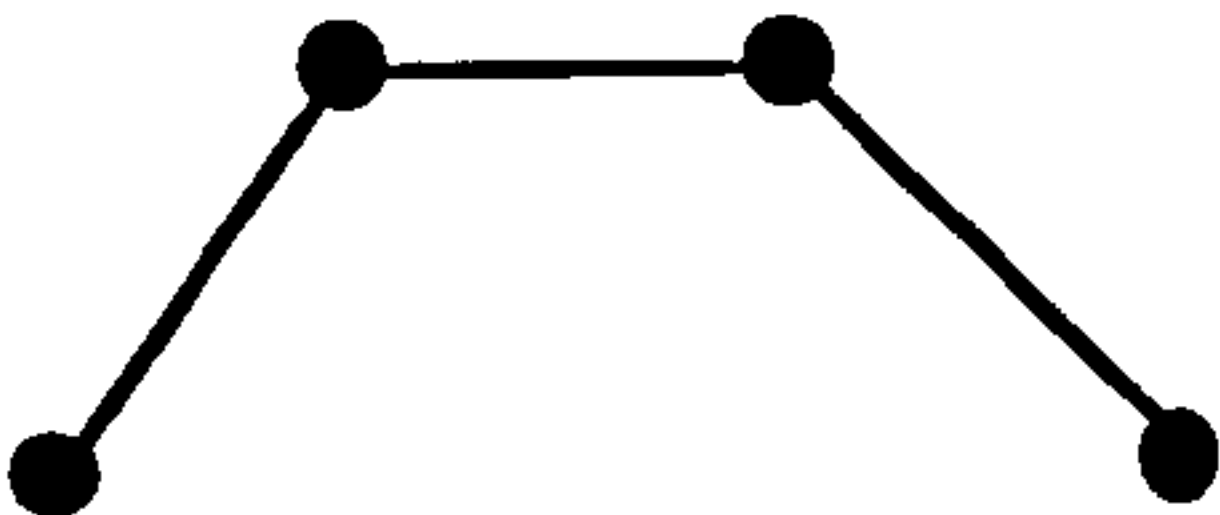
Figure 4.2. Pitch contour models of words with a voiceless vowel and with only voiced vowels (from N. Hattori, 1989: 83-85)

(1) 4 mora words

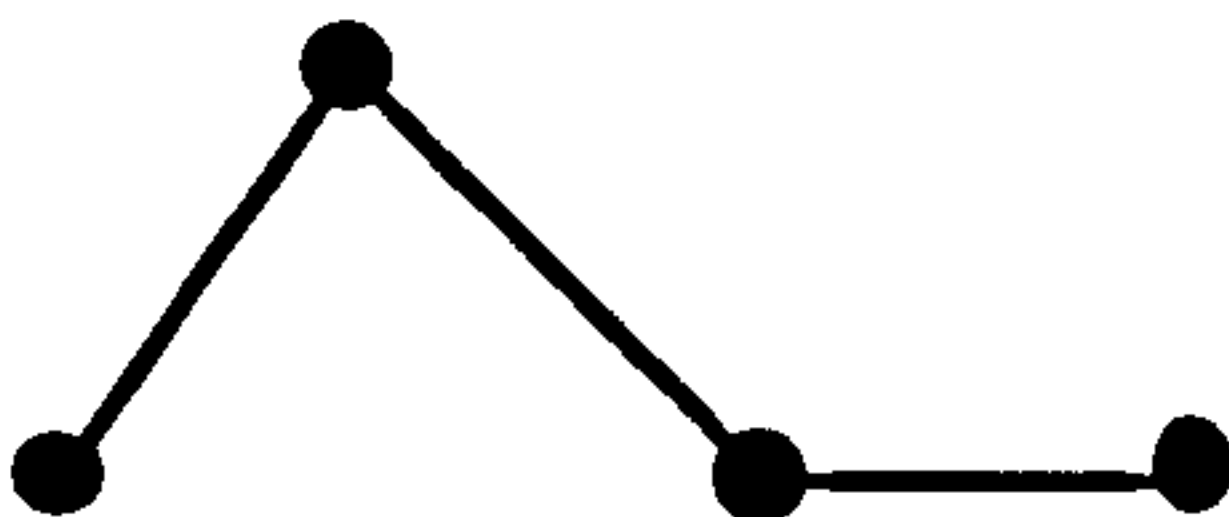
/ho.o.ti̥.ki/<sup>12</sup> 'fire alarm'



- (a) /ha.N.bu̥.N/ 'half'  
/ko.o.i̥.do/ 'latitude'  
/ke.i.ri̥.bu/ 'accountant's department'

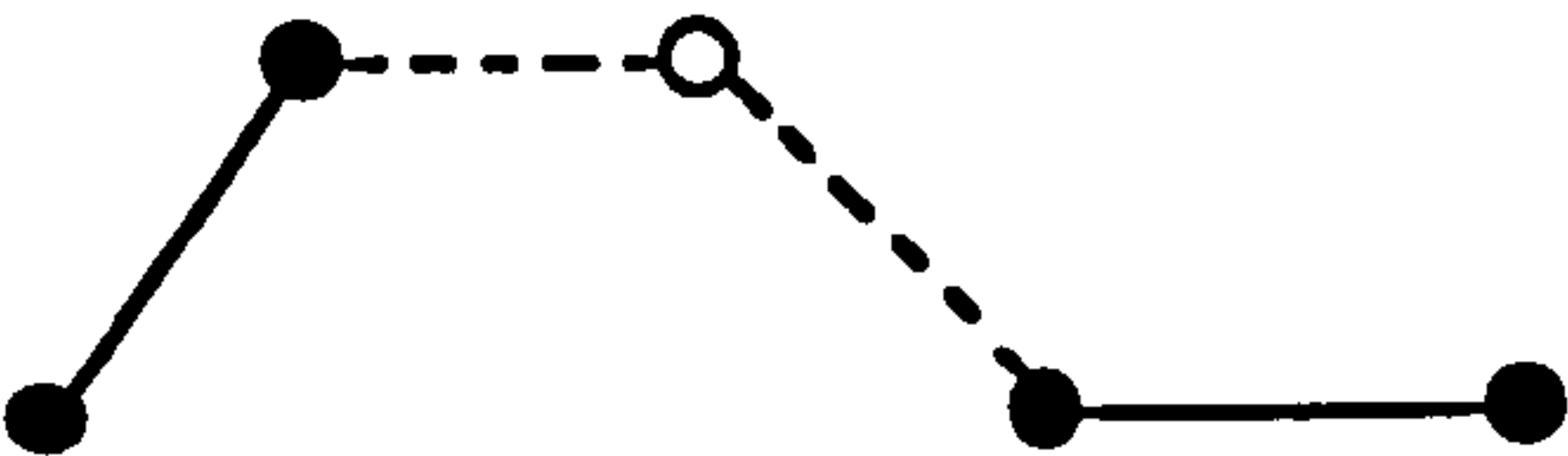


- (b) /hi.mḁ.wa.ri/ 'sunflower'  
/zi.mu̥.i.N/ 'clerk'  
/no.mi̥.mo.no/ 'beverage'

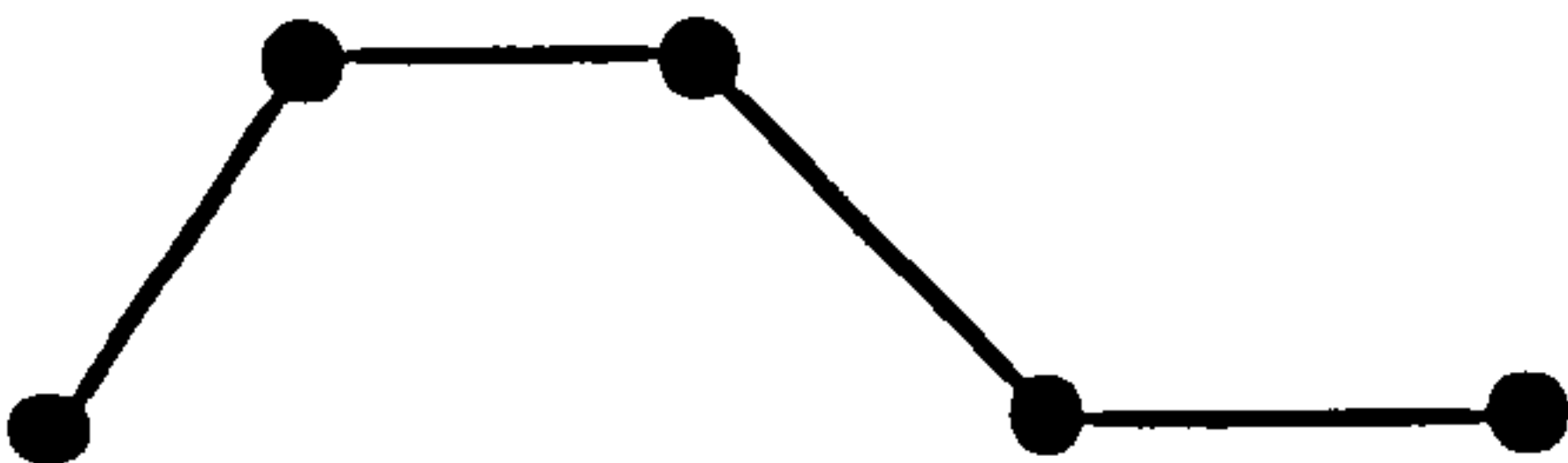


(2) 5 mora words

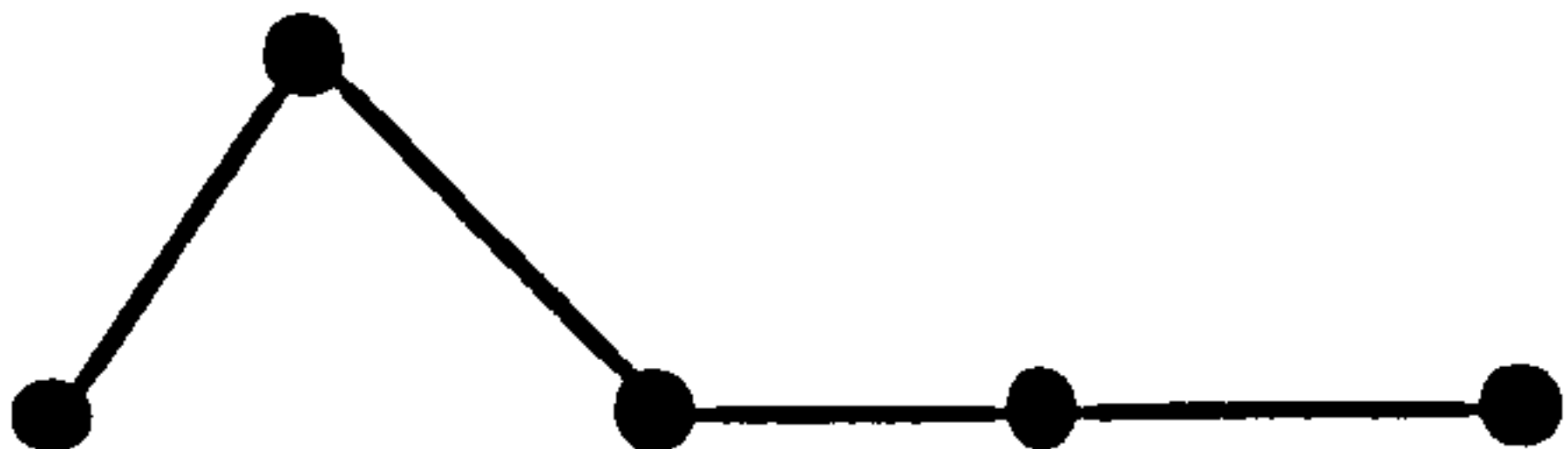
/zi.mu.ki̥.te.i/ 'business regulations'



- (a) /ba.re.ri̥.i.na/ 'ballerina'  
/ze.mi.nḁ.a.ru/ 'seminar'



- (b) /bu.ḁ.i.so.o/ 'unsociable'  
/go.jo̥.o.zi.N/ 'Be aware of ---'



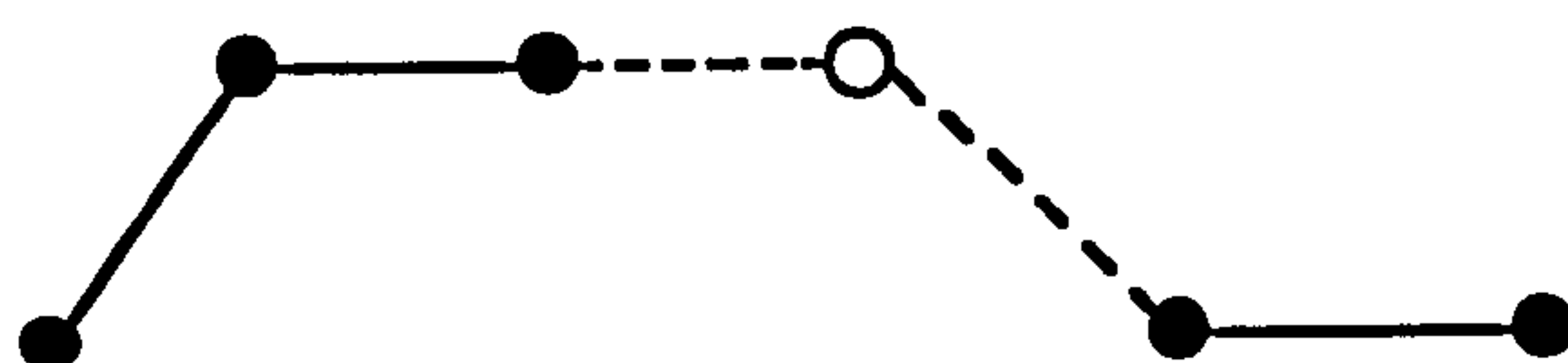
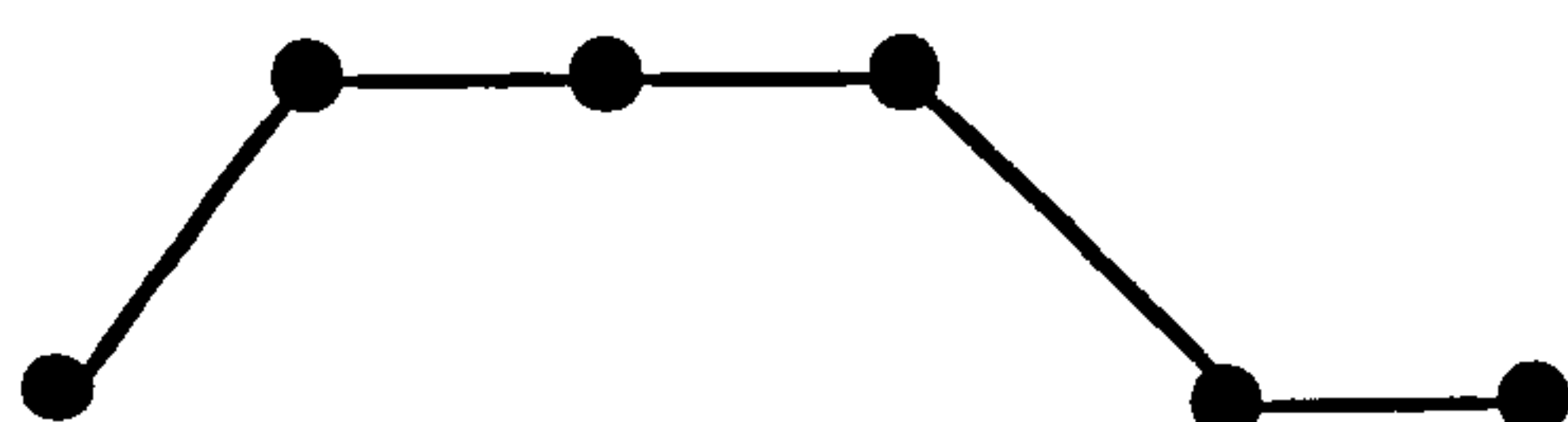
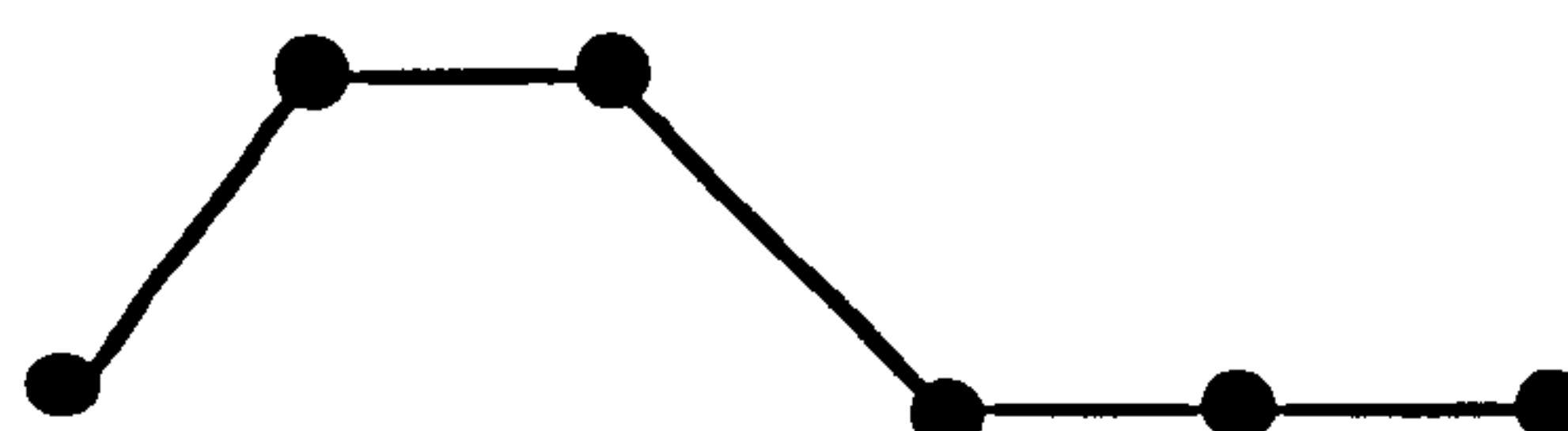
(Figure 4.2 continues)

<sup>12</sup> Potentially devoiced vowels are marked with a symbol /<sub>o</sub>/ underneath the vowels in phonemic transcription for convenience.



(Figure 4.2 continued)

## (3) 6 mora words

*/i.rjo.o.ki̥'.ka.i/* 'medical instruments'(a) */mo.no.no.a'.wa.re/* 'sentiments'*/a.N.mo.ni̥'.u.mu/* 'ammonium'(b) */sa.ra.ri̥'.i.ma.N/* 'office worker'*/ju.u.me.i.zi.N/* 'celebrity'

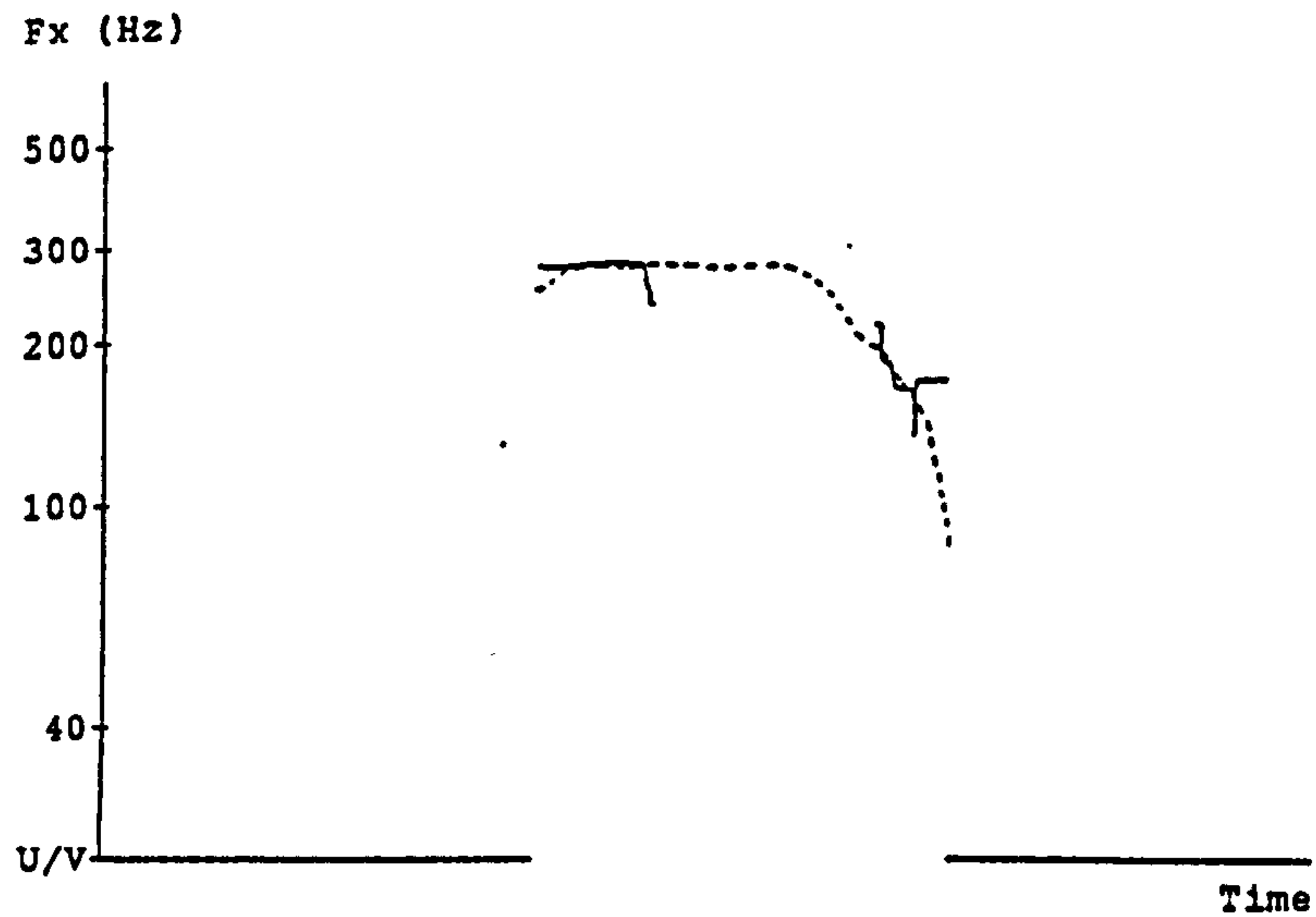
N. Hattori demonstrated that the pitch contours of Figures 4.2 (1) */ho.o.ti̥'.ki/*, 4.2 (2) */zi.mu.ki̥'.te.i/* and 4.2 (3) */i.rjo.o.ki̥'.ka.i/* were rather like that of Figures 4.2 (1a), 4.2 (2a), 4.2 (3a) respectively, although it was not always easy to determine because the vowel was sometimes devoiced too much to be able to differentiate between the (a) and (b) patterns.

Figures 4.3a and 4.3b show examples of superimposition of pitch contours of 4 mora words. The solid line shows a pitch contour of a word */ho.o.ti̥'.ki/* [ho:tɕi̥'ki] 'fire alarm' with a gap for voiceless segments. The broken line in Figure 4.3a shows a model pitch contour of 4 mora words with the accent on their third mora listed in Figure 4.2 (1a) above created by averaging pitch contours of two utterances of each word by six subjects. The broken line in Figure 4.3b shows a model pitch contour of another set of 4 mora words with the accent on their second mora listed in Figure 4.2 (1b) above created by the same method as the model pitch contour in Figure 4.3a. The pitch contour of [ho:tɕi̥'ki] with a gap for voiceless [j] matches the model pitch contour of a word with the accent on the third mora (Figure 4.2 (1a) type) as shown in Figure 4.3a more than that of a word with the accent on the second mora (Figure 4.2 (1b) type) as shown in Figure 4.3b.

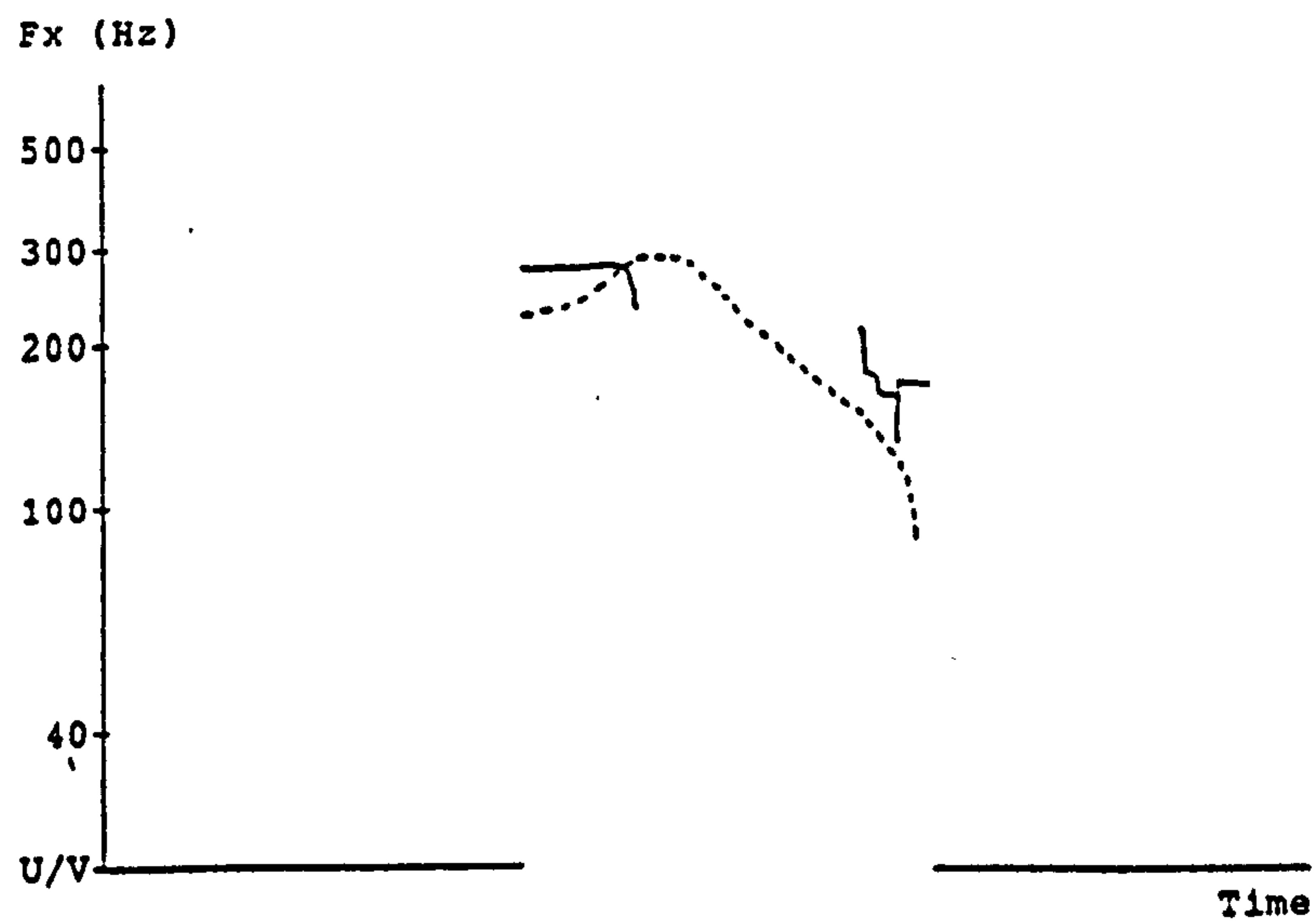
**Figures 4.3a and 4.3b**

**Superimposition of a 'pitch contour of [ho:ti'ki] (the solid line) on the model pitch contours (the broken lines) of four mora words with (a) an accent on the third mora and (b) an accent on the second mora (from N. Hattori, 1989: 93)**

(a)



(b)



N. Hattori (1989: 116) commented that "there was a fixed relationship between a pitch contour and a string of segments of a given number of morae, so that even if part of the pitch contour was missing, it would be still possible to determine or at least classify it from the overall pitch contour". She concluded that sequences containing devoiced vowels seemed to still contain sufficient information to determine the accent location, and that a listener did not work out the accent location solely from auditory information of a particular mora or segment, but rather taking into account the overall pitch pattern of words.

One question here is why a listener can hear a devoiced vowel as accented. When a phonologically accented vowel is devoiced, there is no phonetic evidence of accentedness, i.e. /LHH<sup>1</sup>L/ is realised as [LH?L]<sup>13</sup>. However it is in fact possible to perceive pitch change in a devoiced mora. According to Sugito and Hirose (1988), the perceptual cue for detecting an accent on a devoiced vowel seems to be not actual high pitch in the theoretically accented voiceless vowel and fall of pitch between the devoiced vowel and the following vowel, but the fact that the fundamental frequency of the vowel in the following vowel starts higher than usual and then falls very sharply.

"--- the fundamental frequency (F0) contours of the [o] of [haçj'to] 'the bridge and' with an accented and devoiced preceding vowel have sharp descending tones while those of [haçjto] 'the edge and' have level tones. The [o] vowels of [haçj'to] and [haçjto] were exchanged in a splicing experiment, and the results of hearing tests using these words showed that words with descending third vowels were heard to be accented, while words with level tones that words with descending third vowels were heard as unaccented. These results suggest that it is the following descending tones that make preceding devoiced vowels heard as accented (Sugito and Hirose, 1988: 19).

Sugito and Hirose also found a positive correlation between the pattern of the cricothyroid muscle activity and the raising and lowering of F0. In addition, the activation of the sternohyoid muscle seemed to correlate with a descent in the F0 contours. The onset of F0 lowering was found to be following the cricothyroid peak and a little before the sternohyoid peak. It was also noted that the sternohyoid activity increased prior to the production of words of the low-start accent type. (Sugito and

---

<sup>13</sup> Here "L" means "low pitch" and "H" means "high pitch".

Hirose, 1988: 20-21). The same sharp pitch drop in the following vowel also seems to be the major cue to determine the accent location in the consecutively devoiced sequences (Maekawa, 1991).

Finally, another important point with regard to pitch is that devoicing does not occur in situations with rising intonation where the question particle /-ka/ is omitted.

- |                      |                     |
|----------------------|---------------------|
| (13) [koko ni imasu] | (I stay here.)      |
| [koko ni imasu ka]   | (Do you stay here?) |
| [koko ni imasu] ↗    | (You stay here?)    |

### 4.5.3 Position in an utterance

The effect of mora position in a word on vowel devoicing has not been very extensively studied except for word-final position. Most studies agree that a pause affects devoicing in the same way as a voiceless consonant, therefore a high vowel can be devoiced word finally (see Section 4.5.1.4). However, the results of many experiments suggest that the devoicing rate in word-final position is lower than in other positions. Sakurai (1985: 129) pointed out exceptions to the devoicing rule II (see Section 4.5) where vowel devoicing did not occur even at the end of a word regardless of accent, e.g. the final /u/ in /supa'iku/ 'spike' [sɯpa'iku], the final /i/ in /subekara'si/ 'by all means' [sɯbekara'çi], the final /i/ in /taraima'wasi/ 'passing things from one to another' [taraima'waçi].

Both studies by Maekawa (1989: 144) showed a fairly low devoicing rate in word-final position: about 5% for /ki/ and about 10% for /si/.

Takeda and Kuwabara (1987: 105) analysed /i/-devoicing in the /si/ mora in isolated words and concluded that there was only one devoiced /i/ in word final position out of 82 samples analysed. Word initial devoicing was greater than word medial devoicing:



(14)

initial	>	medial	>	final
(70/224 samples [31%])		(45/188 samples [24%])		(1/82 samples [1%])

Yoshida and Sagisaka (1990) reported slightly different results. They also analysed isolated words and found that the number of devoiced vowels was highest in word-medial position, while the numbers in word-initial and word-final positions were much lower. The number in word-initial devoicing was the lowest in all of their subjects. However, it is dangerous to conclude from these results that devoicing is more likely to occur in the middle of a word than in word-initial or word-final position, because they did not provide the devoicing rate, nor the total number of analysed morae for each position. Moreover, they did not classify their data by the number of morae in a word. The number of morae in a word seems to have some effect on vowel devoicing (see Section 4.5.4), but they simply classified devoicing by its position in terms of the mora number from the end of a word and did not consider the total number of morae in a word; for instance, the vowels /i/ in /ki/ in the words /ki.si/ 'shore' (no accent) and /si.N.bu.N.ki'.sja/ 'newspaper journalist' are both in the second mora from the end, but /kisi/ is a 2 mora word and /siNbuNki'sja/ is a 6 mora word. Therefore, their results are not very reliable apart from the devoicing in word-final position. It should be noted that one of their subjects showed extremely high devoicing in word-final position compared to the others. This suggests that there are large individual differences in word-final vowel devoicing.

In her 1976 paper, Lovins also studied the relationship between devoicing and position in an utterance. In addition to pitch contrast in the test words mentioned in Section 4.5.2, the experiment contrasted positions in a sentence for the same words: e.g./kutu' to kusi' to dottiga joi desuka/ 'Which is better, the shoes or the comb?', and /kusi' to kutu' to dottiga joi desuka/ 'Which is better, the comb or the shoes?'. She concluded that devoicing of accented vowels occurred towards the beginning of an utterance rather than later. She suggested that this was related to reduced subglottal pressure which would make devoicing a mora natural. If Lovins's analysis is correct, the results of Takeda and Kuwabara (1987) seem plausible. Yoshida and Sagisaka

(1990) did not find significant differences in position in a breath group, which contradicts Lovins's account. However, other factors such as the type of adjacent consonants and accent position must be taken into account in combination with the position in a word. Other factors may have a stronger influence on devoicing than the position in a word, but this has not been widely investigated.

#### 4.5.4 Word boundary

Vowel devoicing occurs across a word boundary, or rather it should be said that the type of consonant affects devoicing across a word boundary.

The examples in (15) show the effect of a following consonant on vowel devoicing across a word boundary.

(15)		(a)		(b)
(i)	<u>watashi</u> /watasi/	: [wataɕi - to]	vs	[wataɕi - ɲa] <sup>14</sup>
	'I	'I - and'		'I - subj-marker'
(ii)	<u>tochi</u> /toti/	: [totɕi - to]	vs	[totɕi - ɲa]
	'land'	'land - and'		'land - subj-marker'
(iii)	<u>kisetsu</u> /kise'tu/	: [ki'setsu - to]	vs	[ki'setsu - ɲa]
	'season'	'season - and'		'season - subj-marker'
(iv)	<u>ongaku</u> /o'Ngaku/	: [o'ŋ:aku - to]	vs	[o'ŋ:aku - ɲa]
	'music'	'music - and'		'music - subj-marker'

The last vowels in words (i) to (iv) are high vowels /i/ or /u/, and are preceded by voiceless consonants. When the particle /to/ immediately follows a word as in (a), the last high vowel of each word becomes voiceless as it is surrounded by voiceless consonants. When the subject marker -/ga/ comes immediately after a word as in (b), however, the last vowel of each word is no longer in a devoicing environment, thus it remains voiced. In cases when a speaker inserts a pause between a word and /to/ in (a), the last vowel of the word is likely to be devoiced since it is in the devoicing

<sup>14</sup> Word-internal /g/ is phonetically realised as either [ŋ] or [g] (see Chapter 2 for further discussion).

environment, preceded by a voiceless vowel and followed by a pause (see Section 4.5.1.4).

Fukai (1979) analysed devoicing across an internal word boundary in compound words. He said that when a speaker felt a word juncture in a compound word, a devoiceable vowel immediately before an internal word boundary might remain voiced. He gave an example of the last /i/ in /ikuki/ in a sentence "*Kimi wa hontooni ikuki ka?* 'Do you really intend to go?'. A word ikuki 'an intention to go' consists of two words: iku 'to go' and ki 'intention'. A potentially devoiceable /u/ in /ikuki/ remains voiced, because of an underlying internal word boundary. Sakurai (1985: 129) also listed some examples of voiced high vowels before an internal word boundary in the devoicing environment.

- (16) [ɯ] in kyooakuhannin [kjo:akɯ + hannin] 'brutal criminal'  
 [i] in booe kisuijun [bo:eki + swidʒɯn] 'trade level'

Whether a speaker devoices a vowel or not seems to be dependent on whether a speaker feels a presence of a word boundary during speech, and also dependent on other factors such as speech rate and speech style.

#### 4.5.5 Length of an utterance

The effect of utterance duration has not been widely studied either. This is inevitable as the majority of studies on vowel devoicing are based on data obtained from isolated words, and studies of reading style or connected speech have been very few. Some results suggest that the longer the length of an utterance, the higher is the rate of devoicing. Kuwabara and Takeda (1988: 10) stated that the devoicing rate of /u/ in /hu/ was higher in utterances of four morae or more, than in those of three morae or less. Yoshida and Sagisaka (1990) also studied the relationship between the length of utterance and devoicing by analysing devoicing in the utterance of five morae or more



and of four morae or less<sup>15</sup>, and confirmed the claim of Kuwabara and Takeda. The results of Yoshida and Sagisaka showed that devoicing was less likely in utterances of less than five morae. They suggested that the reason for this was that it would cause difficulty in recognising words if a vowel was devoiced in a short utterance.

#### 4.5.6 Speech style

This factor has not also been well studied, but many linguists think that it is likely that more devoicing occurs in reading and spontaneous speech than in isolated words, and more in casual speech than in speech in a formal style. This tendency may be more salient in the case of non-high vowel devoicing, but high vowel devoicing occurs even in careful or formal speech, and is rather an obligatory process in certain environments (Maekawa, 1989: 143, see Sections 4.4 and 4.5.1.4). The results of Yoshida and Sagisaka (1990) showed higher devoicing rates in reading style than in isolated words. Their subjects devoiced in 205-211 samples out of 457 samples in text reading, which was a devoicing rate of 84-88%, whereas the devoicing rate in isolated words was 75-89%. However, the effect of speaking style or whether the style has any effect on devoicing is not very well known.

#### 4.5.7 Speech rate

The effect of speech rate has not yet been extensively studied, but it is known that speech rate has a great influence on the production of sounds. Research has been carried out concerning the segmental duration, coarticulation, the movement of the articulatory organs (Lehiste, 1970; Gay and Hirose, 1973; Port, 1976; Gay, 1981; Flege, 1988, etc). However, as Maekawa (1989: 47) pointed out, speech rate influences not only the articulation of sounds, but also the physiological mechanisms of the larynx, such as the onset or offset of voicing. With regard to the effect of

---

<sup>15</sup> It was not clearly stated in which category an utterance of five morae belongs to in their study, because they wrote "five morae or more" and "five morae or less" in the literature. I make an assumption that they meant "four morae or less" and "five morae or more".



speaking rate on Japanese vowel devoicing, speaking rate is said to have little connection with high-vowel devoicing because devoicing is expected to occur in certain environments regardless of the speech rate, or speech style (see Section 4.5.1.4). However, it may have some effect in environments where devoicing is less likely to occur; for example, when the vowel is adjacent to a voiced consonant, in an accent nucleus position, non-high vowel devoicing, etc. Even in typical devoicing environments, there may be some influence of speech rate on acoustic or physiological aspects of speech which are not simply manifested in devoicing rates, such as the duration of vowels or morae, weakening of vowels, degree of coarticulation between a vowel and adjacent consonants, etc.

Han (1962a: 92) stated that in the sequence of /Ç<sub>1</sub>V<sub>1</sub>Ç<sub>2</sub>V<sub>2</sub>Ç<sub>3</sub>V<sub>3</sub>/ where all /Ç/s are fricatives, the vowels are reduced to extremely short duration or almost unrecognisable at a fast tempo. She explained the process where /huhuku/ 'discontent' is usually pronounced [ɸɯɸɸkɯ], but at fast tempo it is further reduced to [ɸ:ɸ:kɯ] or [ɸ::kɯ]. Similarly /susuki/ [sɯsɯki] or [sɯsɯki] become [s:s:ki] then [s::ki] as the tempo increases. In fact, it is not possible to separate the sequences of voiceless fricatives as they are just a long continuation of the same fricative noise. However, the contrasts between /huku/ 'clothes' and /huhuku/ 'discontent' and /suku/ 'to become empty' and /susuki/ 'pampas grass' are maintained because of the durational differences of the fricatives. Her observation also suggests that although devoicing is not just a fast speech phenomenon, the effect of speech rate can be manifested in the durations of segments.

Detailed account of the effect of speech rate on vowel devoicing will be discussed in Chapter 6.

#### 4.5.8 Devoicing in consecutive morae

It has been reported that when vowels in two or more consecutive morae are in the devoicing environment some of the vowels remain voiced in order to avoid ambiguity

in recognising words. For example, Han (1962a: 91) stated that in normal speech the devoicing of vowels in two consecutive syllables was rare and devoicing in three successive syllables did not occur. Sakurai (1985: 129) listed two rules for consecutive vowel devoicing: (1) the vowel in one of the devoiceable morae remains voiced when two consecutive morae have potentially devoiceable vowels, e.g. the second /i/ in [kiki'kata] (no accent) 'a way to listen', /i/ in [takitsuke'ru] 'to fire', the second /i/ in [rekijiteki] (no accent) 'historical'; (2) the devoiceable vowel in the middle mora will remain voiced when more than three consecutive morae have potentially devoiceable vowels, e.g. the second /i/ in [kikisute'ru] 'to ignore', and [kikitsuke'ru] 'to hear/learn'.

However, in natural speech it is possible for all the vowels in consecutive morae to be devoiced. It depends on individual speakers, speech rate, speech style, etc., but some patterns of devoicing seem to exist. Maekawa (1989: 148) studied the devoicing patterns for the phrase /sukikirai/ 'like and dislike' in 29 subjects. The results were: [sukiki'rai] (3 people), [sukiki'rai] (19 people), [sukiki'rai] (4 people) and [sukiki:i'rai] (3 people). In another study, Maekawa controlled accent positions (1990). He found that two of his subjects pronounced the word /husitsu/ (usually not accented) with all vowels devoiced in all their utterances ([ɸʉʑitsu]), and the other subject devoiced the second and the third vowels in most of his utterances ([ɸʉʑitsu]). The third subject devoiced all three vowels three times in /husi'tu/ and once in /husitu'/. His results suggested that in the consecutive devoicing environment, vowels preceded by fricatives tend to be devoiced more than vowels preceded by plosives, consecutive devoicing over two morae is possible, and consecutive devoicing even over three morae occurs in natural speech. It does not seem to be very common for all vowels in the consecutive devoicing environment to remain voiced. However, the patterns of voicing and devoicing in the consecutive devoicing environment have not been extensively studied.

According to Vance (1987: 52), both Martin (1952) and Hirayama (1960) denied consecutive devoicing. Their analysis suggested that in consecutive devoicing environments an accented vowel remained voiced and its neighbouring vowels could

be voiceless so long as the combination did not create consecutively devoiced morae. For example, in chishiki [tɕi'ɕiki] 'knowledge' the accented first /i/ remains voiced, then the second /i/ becomes voiceless and the third /i/ must also remain voiced to avoid consecutive devoicing. On the other hand, in tsukusu /tuku'su/ 'to dedicate' the accented second syllable remains voiced, and both the first and third syllables can be devoiced without producing consecutively devoiced morae [tsuɰkɯ'su]. When none of the morae in the sequence is accented, there are usually some variations but the first of the sequence is often likely to be devoiced.

Han's study (1962a) showed that vowels preceded by fricatives were more often devoiced than those preceded by plosives. However, this particular result was not very convincing. She compared pronunciations of the sentences /sukusukuto nobita/ 'Grew fast' and /kusukusu to waratta/ 'giggled'. The spectrograms of the utterances she presented showed devoiced [ɯ] after [s] as [suɰkɯ'suɰkɯ] and [kɯ'suɰkɯsuɰ]. The problem is that accent was also involved here. NHK Accent dictionary (1985) lists both [suɰ'kɯsuɰkɯ] and [suɰkɯsuɰkɯ] (no accent), and also [kɯ'suɰkɯsuɰ] and [kɯsuɰkɯsuɰ] (no accent), where vowels after fricatives are not necessarily devoiced. However, if the vowel following a fricative is accented it remains voiced, but a vowel in an adjacent mora, preceded by a plosive, undergoes devoicing. It is therefore not clear whether Han was justified in proposing an independent principle depending on the preceding consonant.

Han (1962a) also stated that /u/ was more likely to devoice than /i/ if both are involved in a sequence of the type under discussion, i.e. consecutive devoicing environments. Her example was /pikupiku to/ (twitch-twitch), which she said was more often pronounced [pikuɰpiku to] than [pɰkɯpɰkɯ to], and deduced that this was due to the shorter inherent duration of /u/ (see. Section 4.5.1.1). However, this observation must be re-examined as she did not consider the accent location for the word, and its accent position is normally on the first mora, thus /pi'kupiku/. Her observation that [pikuɰpiku to] is more common than [pɰkɯpɰkɯ to] is simply because a speaker avoids devoicing on an accented vowel. The Accent Dictionary lists [pi'kɯpɰkɯ] and [pɰkɯ'pɰkɯ] both with a voiceless [j] regardless of the accent position.



The devoicing in consecutive morae seems to be influenced greatly by a combination of various devoicing factors, and how these factors interact each other and influence on devoicing should be examined in detail.

## 4.6 Conclusions

Although there has been extensive research on various aspects of vowel devoicing in Japanese, there are still many aspects which still require further study. The main focus of these studies has been the phonological environment, especially the effects of preceding and following consonants, pitch and accentedness, and the position in an utterance. These factors have been investigated on an individual basis, as they are easy to quantify by measuring devoicing rates. However, the effect of these factors in combination has not been so well studied. For instance, the presence of one blocking factor may be enough to block devoicing in one environment, but more than one factor may be required in another environment. In one environment, phonetic factors such as the type of preceding consonant or speaking tempo can dominate, promoting devoicing and overcoming the presence of blocking factors. However, in another environment, despite ideal phonetic conditions (for promoting devoicing), phonological factors such as accentedness or positions in an utterance can dominate and block devoicing process.

Experimental conditions in these studies varied enormously: from one subject to 29 subjects, and from using only a few sentences to a large scale speech database. One common feature of these studies is a categorical distinction of voiceable vowels: namely voiced or voiceless. The gradient nature of vowel devoicing has never been discussed, nor the degree of devoicing has never been quantified.

The results suggested that vowel devoicing was generally an almost regular process, though it was not categorical. However, the effect of proposed blocking factors varied from study to study. The results of the studies sometimes showed very



high devoicing rates despite the presence of a blocking factor, and sometimes showed very low devoicing rates in an ideal environment for devoicing. One of the conditions in which devoicing is not always compulsory is the consecutive devoicing environment. All studies agreed that not all vowels are devoiced in the consecutive devoicing environment. This might mean that with ideal environments, vowel devoicing may sometimes fail to occur in consecutive devoicing environments.

Consecutive devoicing is one of the areas which has not been well studied. This is an important aspect worthy of further investigation, by looking at vowel devoicing in two separate conditions, namely single and consecutive devoicing environments. The effect of blocking factors could vary in these two environments.

In the later chapters of this thesis, I will take the single and consecutive devoicing environments as the main condition which has a major controlling effect on devoicing, and study this with relation to other suggested phonetic factors (speech rate and type of preceding consonant) and phonological factors (presence of accent and following internal word boundary), as well as the combined effect of these factors.

## Chapter 5

# The effect of blocking factors and constraints on vowel devoicing

### 5.1 Introduction

As discussed in Chapter 4, previous studies have shown that there are many factors involved in Japanese vowel devoicing. Factors affecting vowel devoicing suggested in the literature are (a) the vowel itself, (b) the preceding consonant, (c) the following consonant, (d) the vowel in the following mora, (e) pitch and accent, (f) position in an utterance, (g) length of an utterance, (h) speech style, (i) speech rate, (j) presence of an internal word boundary, (k) regional differences, and (l) generational difference. The effects of some individual factors have been studied extensively (Han, 1962a, Kuriyagawa and Sawashima, 1984, 1985, 1986, 1987 and 1989, Takeda and Kuwabara, 1987, Yoshida and Sagisaka, 1990, etc.) (for details, see Chapter 4). There is general agreement on some of the factors involved, but others are still disputed or not yet fully investigated.

Among phonetic and phonological factors, all the literature agrees on the required phonological environments for vowel devoicing in Standard Japanese, i.e. when the high vowels /i/ and /u/ occur between voiceless consonants, or between a voiceless consonant and a pause at the end of an utterance (Han, 1962a, Sakurai, 1985, Vance, 1987, Maekawa, 1989, etc.). With regard to the position in an utterance, many studies agree that the devoicing rate in an utterance final position is low, and utterance final devoicing is often associated with certain lexical items, such as *-desu*

'COPULA (POLITE)', *-masu* 'POLITE' especially in a sentence final position with falling intonation (Lovins, 1976, Sakurai, 1985, Takeda and Kuwabara, 1987, Maekawa, 1989, Yoshida and Sagisaka, 1990, etc.).

As regards the effect of accent on vowel devoicing, recent research has revealed that a vowel can be devoiced even in an accented position. The phenomenon of accent shift to a preceding mora, in order to avoid devoicing on an accented vowel, seems to be obsolescent. N. Hattori (1989) showed that when an accented vowel of a word was devoiced, the overall pitch pattern of the word retained the same pattern as a word with the same number of morae with the accent on the same position. Sugito and Hirose (1988) showed that a devoiced vowel was still perceived as accented and the perceptual cue to the accent lay not on the vowel itself but on the following mora: there was a rapid fall of fundamental frequency at the beginning of the following mora.

As for the length of an utterance, devoicing seems to be less common in short utterances because devoicing vowels makes the short utterance unclear, as it is difficult to guess the word from the information provided by other words (Yoshida and Sagisaka, 1990).

The effects of other factors listed above have not been fully investigated. For instance, differences in speech style and speaking rate are not, in fact, very well investigated, although there has been a strong belief that devoicing occurs more often in casual speech and at faster tempo (Hasegawa, 1977: 127, Shibatani, 1987: 137, etc). The effect of an internal word boundary and the type of a vowel in the following mora has also not been well investigated.

The effects of some factors have been studied extensively, but the results have not been universally agreed. For example, with respect to the type of preceding and following consonants, most studies rated vowel devoicing higher when preceded by fricatives than when preceded by other types of consonants, namely plosives and affricates, but as between plosives and affricates, the order of effective influence is

disputable. The effect of the type of following consonants and the combination between preceding consonants and following consonants has not been well investigated. As for the position in an utterance, except for utterance final position, the different studies reported different results concerning the effect of different positions.

However, it is usually the case in vowel devoicing that several factors are involved at the same time and possibly interact with each other. But, very little research has been carried out in the situations when more than one factor exists in the mora in question and in cases where consecutive devoicing sites occur. For example, although the presence of accent does not seem to affect devoicing, what would happen if it coincided with an internal word boundary? In the consecutive devoicing environment, do the proposed blocking factors have the same effects as in single devoicing sites? If there is more than one devoicing factor involved, which factor has the greatest effect?

In this chapter, I will examine the effect of some of the suggested phonetic and phonological devoicing factors, and how devoicing in phonetically ideal environments is constrained by phonological factors. The phonetic factor chosen is the type of preceding consonant. The phonological factors chosen are accentuation, presence of a following internal word boundary and devoicing conditions, namely the single and consecutive devoicing environments. Devoicing rates will be analysed by (i) the type of the preceding consonant, (ii) accentuation, and (iii) presence of an internal word boundary, in the citation form of words with (a) a single devoicing site and (b) with consecutive devoicing sites. The interaction between these factors will also be examined.

## 5.2 Experimental hypotheses

Vowel devoicing is considered to be a fairly common phenomenon, but most studies agree that the devoicing process sometimes fails to occur in the consecutive



devoicing environment. Therefore, it is necessary to consider vowel devoicing in the two separate environments: i.e. the single and consecutive devoicing environments. A certain blocking factor may be effective in the consecutive environment, but have little effect in the single environment.

As mentioned in Chapter 4, Section 4.5.8, Han (1962a) studied the combination of several factors in consecutive devoicing environments. She showed that a vowel preceded by a fricative was devoiced to a greater extent than a vowel preceded by a plosive. However, in all the test words, vowels preceded by fricatives were not accented but vowels preceded by plosives were. Therefore, the results may actually have been due to accentuation.

Vance (1987: 52) stated that in the case of potential consecutive devoicing, the accented vowel must remain voiced to avoid consecutive devoicing. Sakurai (1985: 129) proposed the following rules for consecutive devoicing: (1) in the case of two potential consecutive devoicing sites, the second potentially devoiceable vowel remains voiced, and (2) the vowel in the middle mora should remain voiced when three or more consecutive morae have potentially devoiceable vowels. Both Vance and Sakurai denied the possibility of consecutive devoicing, but none of their rules can provide all possible combinations of blocking factors for devoicing.

Vance's rule works for two consecutive devoicing sites. For example, in the sequence<sup>1</sup> /C<sub>1</sub>V<sub>1</sub><sup>ˈ</sup>C<sub>2</sub>V<sub>2</sub>/, /V<sub>1</sub>/ remains voiced and /V<sub>2</sub>/ becomes voiceless, and in the case of /C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>2</sub><sup>ˈ</sup>/, V<sub>2</sub> remains voiced and V<sub>1</sub> undergoes the devoicing process. However, in the case of consecutive devoicing sites over three morae, such as /C<sub>1</sub>V<sub>1</sub><sup>ˈ</sup>C<sub>2</sub>V<sub>2</sub>C<sub>3</sub>V<sub>3</sub>/, in which all the vowels /V<sub>1</sub>/, /V<sub>2</sub>/, and /V<sub>3</sub>/ are potentially devoiceable and the vowel /V<sub>1</sub>/ in the first mora is accented, Vance's rule leaves /V<sub>1</sub>/ voiced and makes /V<sub>2</sub>/ and /V<sub>3</sub>/ voiceless. As a result it creates consecutive devoicing of /V<sub>2</sub>/ and /V<sub>3</sub>/, which is contradictory to his statement, as he worked out the rule in order to avoid consecutive devoicing.

---

<sup>1</sup> An accent is marked with a symbol "ˈ": in /C<sub>1</sub>V<sub>1</sub><sup>ˈ</sup>C<sub>2</sub>V<sub>2</sub>/, the vowel V<sub>1</sub> is accented, i.e. there is a pitch drop between /C<sub>1</sub>V<sub>1</sub>/ and /C<sub>2</sub>V<sub>2</sub>/.

Sakurai's rule seems plausible as it manages to avoid consecutive devoicing even in more consecutive sites, such as over four morae. However, his rule still has some problems. For instance, as Maekawa (1989: 143) pointed out, a word *shukufuku* /sjukuhuku/ (no accent) 'blessing' has various voicing/devoicing possibilities, but the pattern [çʷkʷɸʷkʷɸ] definitely sounds strange. Sakurai's rule does not consider the type of adjacent consonants, thus it leaves some factors to be accounted for.

As far as word boundary is concerned, Fukai (1978: 52) and Sakurai (1985: 129) stated that a devoiceable vowel may remain voiced when a speaker feels there is an internal word boundary (see Chapter 4, Section 4.5.5). On the other hand, it is true that devoicing occurs across a word boundary (Sakurai, 1985, Vance, 1987, Maekawa, 1989, etc.). None of the studies have worked out devoicing patterns when a devoiceable vowel is accented and followed by a word boundary. The combination of these factors and the type of preceding consonant may have to be accounted for.

As for the proposed blocking factors, both accentuation and the presence of internal word boundary seem to block devoicing, but their effects are not categorical. These factors do not consistently block devoicing. The combined effect of these blocking factors has not been well studied. The presence of one factor may not be enough to block devoicing, but presence of more than one factor may block devoicing.

Following the results of other studies on the effect of individual factors, I will hypothesise the following for the experiment:

- (i) There is a difference in the effect of blocking factors between single and consecutive devoicing environments; blocking factors are effective in the consecutive devoicing environment, but not effective in the single devoicing environment: i.e. high vowels usually become voiceless in the single devoicing environment even when there are blocking factors.

- (ii) When only one of the blocking factors on vowel devoicing is present in a devoicing site, its effect is minimal, but when there is more than one blocking factor, their combined effect blocks devoicing in the consecutive devoicing environment.

### **5.3 Experimental Methods**

#### **5.3.1 Subjects and procedure**

An experiment was carried out to examine the effects of some of the suggested blocking factors, i.e. (i) the type of preceding consonant, (ii) accentuation, and (iii) presence of an internal word boundary, in two devoicing conditions, namely (a) a single devoicing and (b) consecutive devoicing environments.

Six people, two male (TM and NM) and four female (HM, MT, CO and MO), took part in the experiment. All of them, their age ranging from 27 to 62 years old, were brought up in the Tokyo area and are speakers of Standard Japanese. None of the subjects had much knowledge of phonetics and they did not know the purpose of the experiment.

The subjects read 41 test words listed below in isolation. All recordings were made in an anechoic recording studio in the National Language Research Institute, Tokyo. All data were recorded digitally using a SONY DAT tape recorder. All the recordings were analysed using Waves+ speech analysis on a SUN workstation. The recordings were digitised at a sampling rate of 16 KHz.

#### **5.3.2 Materials**

##### **5.3.2.1 Test words**

Forty-one test words listed in Table 5.1a and 5.1b below were chosen for the experiment. Each test word included at least one voiceable vowel, i.e. /i/ or /u/



surrounded by voiceless consonants, or preceded by a voiceless consonant and followed by a pause. Two experimental conditions are set up: (a) a single devoicing condition and (b) a consecutive devoicing condition. Test words in the single devoicing condition have one devoiceable vowel and without any other devoiceable vowels in directly adjacent morae. Test words in the consecutive devoicing condition have more than one devoiceable vowel in successive morae. 20 of them listed in Table 5.1a include single devoicing sites, e.g. /heik*ī*-ko/ 'arsenal', /tetu-ka'buto/ 'steel helmet' (the vowels written in bold italics are devoiceable), and the remaining 21 words listed in Table 5.1b contain consecutive devoicing sites, e.g. /bokut*ī*ku/ 'livestock farming' and /kut*u*sita-ko'udjou/ 'sock factory', 41 test words containing the total of 74 devoicing sites. Words /masi'kaku/ 'square shape', /nat*ī*ke'ikoku/ 'Nachi Gorge' and /hit*ā*ti-ka'igaN/ 'Hitachi Beach' have two devoicing sites in a word, but the devoiceable vowels are not in morae adjacent to each other. Therefore, the word /hitati-ka'igaN/ does not have a consecutive devoicing site, but two single devoicing sites.

There were no nonsense words included in the test words although some words were not be very common, but the test words did include some loan words. The test words including /pi/ and /pu/ sequences are originally loan words, i.e. /daNpu-kaa/ 'dump truck', /kaputiino/ 'cappuccino', /sjaapu-peNsiru/ 'propelling pencil' and /kasupi-kai/ 'Caspian Sea'.

Some names of landmarks were made up as "existing place name + landmark", but they are very likely to exist, e.g. *Akitsu Kooen* 'Akitsu Park', *Nachi Keikoku* 'Nachi Gorge', etc.. The accent placement was based on The NHK Accent Dictionary.



Table 5.1a Test words with a single devoicing site

	<i>/i/</i>		<i>/u/</i>	
	test word	gloss	test word	gloss
<i>/k/</i>	<i>/heikī-ko/</i> <sup>2</sup>	arsenal	<i>/takaku-ke'iei/</i>	diversification
	<i>/sjokī-kaN/</i>	cabinet secretary	<i>/sigaku-ka'ikaN/</i>	private school hall
			<i>/suizo'ku-kaN/</i>	aquarium
<i>/s/</i>	<i>/ma-sī'kaku/</i>	square shape	<i>/kakujasu-tu'aa/</i>	package holiday
			<i>/garasu-ko'udjou/</i> <sup>3</sup>	glass factory
<i>/sj/</i>			<i>/kousju'-kei/</i>	death by hanging
			<i>/seNsjū-kjo'okai/</i>	athletic association
<i>/t/</i>	<i>/nati-ke'ikoku/</i>	Nachi Gorge	<i>/akatuti-jama/*</i>	Mt. Akatsuchi
	<i>/hitati-ka'igaN/</i> <sup>4</sup>	Hitachi Beach	<i>/keNritu-ko'ukou/</i>	state high school
			<i>/tetu-ka'buto/</i>	steel helmet
<i>/h/</i>	<i>/sjouhi-ke'ikaku/</i>	spending plan		
	<i>/asahi-ka'igaN/</i>	Asahi Beach		
	<i>/hitati-ka'igaN/</i> <sup>3</sup>	Hitachi Beach		
<i>/p/</i>			<i>/daNpu'-kaa/</i>	dump truck
			<i>/kaputi'ino/</i>	cappuccino
			<i>/sjaapu-pe'Nsiru/</i>	propelling pencil

Devoiceable vowels are written in bold italics.

' indicates a preceding vowel is accented.

- indicates an obvious internal word boundary.

\* indicates no accented word

<sup>2</sup> The sequence */ei/* is normally phonetically realised as [e:].

<sup>3</sup> The sequence */ou/* is normally phonetically realised as [o:].

<sup>4</sup> The first and second */i/*s are devoiceable, but both of them are in single devoicing environments.

**Table 5.1b Test words with consecutive devoicing possibilities**

test word	gloss
/dookutu/*	cave
/hihu-ka/*	dermatology unit
/kasupī-kai/	Caspian Sea
/desaki-kīkaN/	district office of the central government
/garasu'-ki/	glassware
/hootī-ki/	fire alarm
/sjoohī-ti/	consumer belt
/kokuhuku/*	restoration
/bokutiku/*	livestock farming
/kiNki-tīhoo/	the Kinki District
/toohoku-tīhoo/	the Toohoku District
/aiti-tīhoo/	the Aichi area
/gihu-tīhoo/	the Gifu area
/kiri hukī-ki/	sprayer
/tjoohu-kīti/	Choofu Base
/jusjutu-kīti/	exporting centre
/huku-to'siN/	the hub of the city
/kutu-ko'odjoo/	shoe factory
/kutusita-ko'odjoo/	sock factory
/tukusi-te'tudoo/	Tsukushi Railway
/rjokuti-ko'oeN/	forest park

Devoiceable vowels are written in bold italics.

- ' indicates a preceding vowel is accented.
- indicates an obvious internal word boundary.
- \* indicates no accented word

### 5.3.2.2 Blocking factors

#### 5.3.2.2.1 Accentuation

Many studies agree that vowel devoicing is less likely to occur on an accented mora; either devoicing is blocked or accent is shifted to an adjacent mora (Han, 1962a, Sakurai, 1985, Takeda and Kuwabara, 1987, etc.). However, it has been widely recognised that some speakers, especially speakers from the younger generation, tend to devoice an accented vowel (see Chapter 4, Section 4.5.2). In order to examine the effect of accent, eight of the test words have an accented high vowel (see Tables 5.2a and 5.2b below).

#### 5.3.2.2.2 Internal word boundary

In order to examine the effect of internal word boundary, quite a few compound words were included in the test words. However, the degree of compoundness varies. Some words have a fairly clear and obvious internal boundary while others are less obvious. For instance, *ryokuchikooen* 'forest park' and *kutsukoojoo* 'shoe factory' have a clear word boundary. Devoiceable vowels followed by an internal word boundary are underlined in this section.

- (1) a. *ryokuchi*            'green area'            +            *kooen*            'park'
- b. *kutsu*                'shoe'                +            *koojoo*            'factory'

On the other hand, words such as *hootiki* 'fire alarm' and *kooshukei* 'death by hanging' are compound words and have a fairly clear internal word boundary as shown in (2), but they are almost always uttered as if they were one word. It is very rare that only the first parts of the words (*houchi* and *koushu*) are used independently.

- (2) a. *hoochi*                'informing'            +            *ki*                'apparatus'

- b. *kooshu*                    'strangling                    +                    *kei*                    'punishment'  
   the neck'

Sometimes words with a clear word boundary are combined with each other in ordinary conversation, e.g. *desakikikan* 'district offices of the central government' and *Toohokuchihoo* 'the Toohoku District':

- (3) a. *desaki*                    'local branch'                    +                    *kikan*                    'office'
- b. *Toohoku*                    'north east'                    +                    *chihoo*                    'district'

A native speaker may not be aware of word boundaries in some words even though strictly speaking they are present, e.g. *bokuchiku* /bokutiku/ 'livestock farming' and *kokufuku* /kokuhuku/ 'restoration'.

- (4) a. *boku*                    'pastral                    +                    *chiku*                    'livestock'
- b. *koku*                    'hardship'                    +                    *fuku*                    'to overcome'

The test words in the groups of (1), (2) and (3) above were treated as words with an internal word boundary, whereas the test words in the group of (3) were treated as words without an internal word boundary.

5.3.2.2.3 Effect of combined blocking factors

The effect of the proposed blocking factors has been studied extensively, but individually. In this experiment, I will examine the effect of combined factors as well as their individual effects on vowel devoicing.

As shown in Tables 5.2a and 5.2b below, some devoiceable vowels (written in bold italics) have only one proposed blocking factor, namely being accented or followed by an internal word boundary, but some have more than one. 8 devoiceable vowels are only accented, e.g. /ma-sĩ'kaku/ 'square shape' and /kiNki-ti'hoo/ 'the Kinki



district'. 25 devoiceable vowels are followed by an internal word boundary only, e.g. /sjaapu-pe`Nsiru/ 'propelling pencil' and /tetu-ka`buto/ 'steel helmet'. There are 9 devoiceable vowels that have the two blocking factors, i.e. are accented and followed by an internal word boundary, e.g. /heiki`-ko/ 'arsenal', /sjoki`-kaN/ 'cabinet secretary', /garasu`-ki/ 'glassware', /kousju`-kei/ 'death by hanging', /houti`-ki/ 'fire alarm', /shouhi`-ti/ 'consumer belt', /daNpu`-kaa/ 'damp truck' and /kirihi`-ki/ 'sprayer'. 20 devoiceable vowels are neither accented nor followed by an internal word boundary, e.g. /dookutu/ 'cave', /akatuti-jama/ 'Mt. Akatsuchi', /kaputi`ino/ 'cappuccino', /kokuhuku/ 'restoration' and /bokutiku/ 'livestock farming'. 12 devoiceable vowels occurred word finally, e.g. /nati-ke`ikoku/ 'Nachi Gorge', /dookutu/ 'cave'.

#### 5.3.2.2.4 Position in an utterance

As discussed in Chapter 4, Section 4.5.3, the position in an utterance is another important factor for the devoicing. There was no agreement on likeliness of devoicing between utterance-initial and utterance-medial, but most studies agree that devoicing is rare in utterance-final position (Sakurai, 1985, Takeda and Kuwabara, 1987, Maekawa, 1989, etc.). Therefore, devoiceable vowels in utterance-final position (a word-final position in this experiment) were avoided in the single devoicing environment. There were three instances of devoiceable vowels in word-final position, namely /ma-si`kaku/, /nati-ke`ikoku/ and /sjouhi-ke`ikaku/, but they were not controlled devoiceable vowels. The chosen test words happened to have word-final devoiceable vowels. However, this is not the case in the consecutive devoicing environment. It was inevitable to use words with word-final devoiceable vowels in the consecutive devoicing environment.

Utterance-final devoiceable vowels occurred in nine test words. The word-final devoiceable vowels were included to obtain devoicing rates. However, acknowledging the tendency that vowel devoicing seemed to be rare in the utterance-final position, separate devoicing rates were calculated: (a) including devoiceable

vowels in word-final position, and (b) excluding devoiceable vowels in word-final position.

Tables 5.2a and 5.2b list devoiceable vowels in the test words grouped by blocking factors and the type of preceding consonants in the single and consecutive devoicing environments. Table 5.2c lists devoiceable vowels in the test words in the utterance-final position.

**Table 5.2a    List of devoiceable vowels in the single devoicing environment by blocking factor and the type of preceding consonant (devoiceable vowels in question are in bold italics)**

blocking factors		type of preceding consonant	test words	
accented	+boundary	plosive	/heik <i>ĩ</i> -ko/	/sjok <i>ĩ</i> -kaN/
			/daNpu'-kaa/	
	-boundary	fricative	/kousju'-kei/	
		fricative	/ma-s <i>ĩ</i> 'kaku/	
not accented	+boundary	plosive	/takaku-ke'iei/	/sigaku-ka'ikaN/
			/suizo'ku-kaN/	/sjaapu-pe'Nsiru/
		fricative	/kakujasu-tu'aa/	/garasu-ko'udjou/
			/seNsju-kjo'okai/	/sjouhi-ke'ikaku/
			/asahi-ka'igaN/	
		affricate	/nati-ke'ikoku/	/hitati-ka'igaN/
			/keNritu-ko'ukou/	/tetu-ka'buto/
	-boundary	plosive	/kaputi'ino/	
		fricative	/hitati-ka'igaN/	
		affricate	/akatuti-jama/	

**Table 5.2b**    **List of devoiceable vowels in the consecutive devoicing environment by blocking factor and the type of preceding consonant**

blocking factors		type of preceding consonant	test words	
accented	+boundary	plosive	/kasupĩ'-kai/	/kiri hukĩ'-ki/
		fricative	/garasu'-ki/	/sjoohĩ'-ti/
		affricate	/hootĩ'-ki/	
	-boundary	plosive	/desaki-kĩ'kaN	/tjoohu-kĩ'ti/
			/jusjutu-kĩ'ti/	
		affricate	/kiNki-tĩ'hoo/	/toohoku-tĩ'hoo/
			/aiti-tĩ'hoo/	/gi hu-tĩ'hoo/
not accented	+boundary	plosive	/desaki-kĩ'kaN/	/kiNki-tĩ'hoo/
			/toohoku-ti'hoo/	/huku-to'siN/
		fricative	/hi hu-ka/	/gi hu-ti'hoo/
			/tjoohu-ki'ti/	/tukusi-te'tudoo/
		affricate	/aiti-ti'hoo/	/jusjutu-ki'ti/
			/kutu-ko'odjoo/	/rjokuti-ko'oeN/
	-boundary	plosive	/dookutu/	/koku huku/
			/bokutiku/	/kutu-ko'odjoo/
			/kutusita-ko'odjoo/	/tukusi-te'tudoo/
			/rjokuti-ko'oeN/	
		fricative	/hi hu-ka/	/kasupi'-kai/
			/kokuhuku/	/kiri huki'-ki/
			/jusjutu-kĩ'ti/	/huku-to'siN/
			/kutusita-ko'odjoo/	
		affricate	/bokutiku/	/tu kusi-te'tudoo/
			/kutusita-ko'odjoo/	

**Table 5.2c** List of test words with a devoiceable vowel in word final position

/ma-si`kaku/	/nati-ke`ikoku/	/sjouhi-ke`ikaku/	/garasu`-ki/
/hooti`-ki/	/kokuhuku/	/bokutiku/	/kiri huki`-ki/
/dookutu/	/sjoohi`-ti/	/tjoohu-ki`ti/	/jusjutu-ki`ti/

### 5.3.2.3 Phonetic environments for the test words and conditions

In order to examine the effect of type of preceding consonants on vowel devoicing, the preceding consonants included all possible consonant types, namely /k, p, s, sj, h, t/. However, the vowels were always followed by plosives, if not in word final position, which enabled easier segmentation for measuring segmental durations for the studies described in Chapter 7. There are many words in which a devoiceable vowel is followed by a voiceless fricative. However, when the high vowel is devoiced in /fricative + V + fricative/ sequences, segmental boundaries are not very clear and therefore difficult to locate.

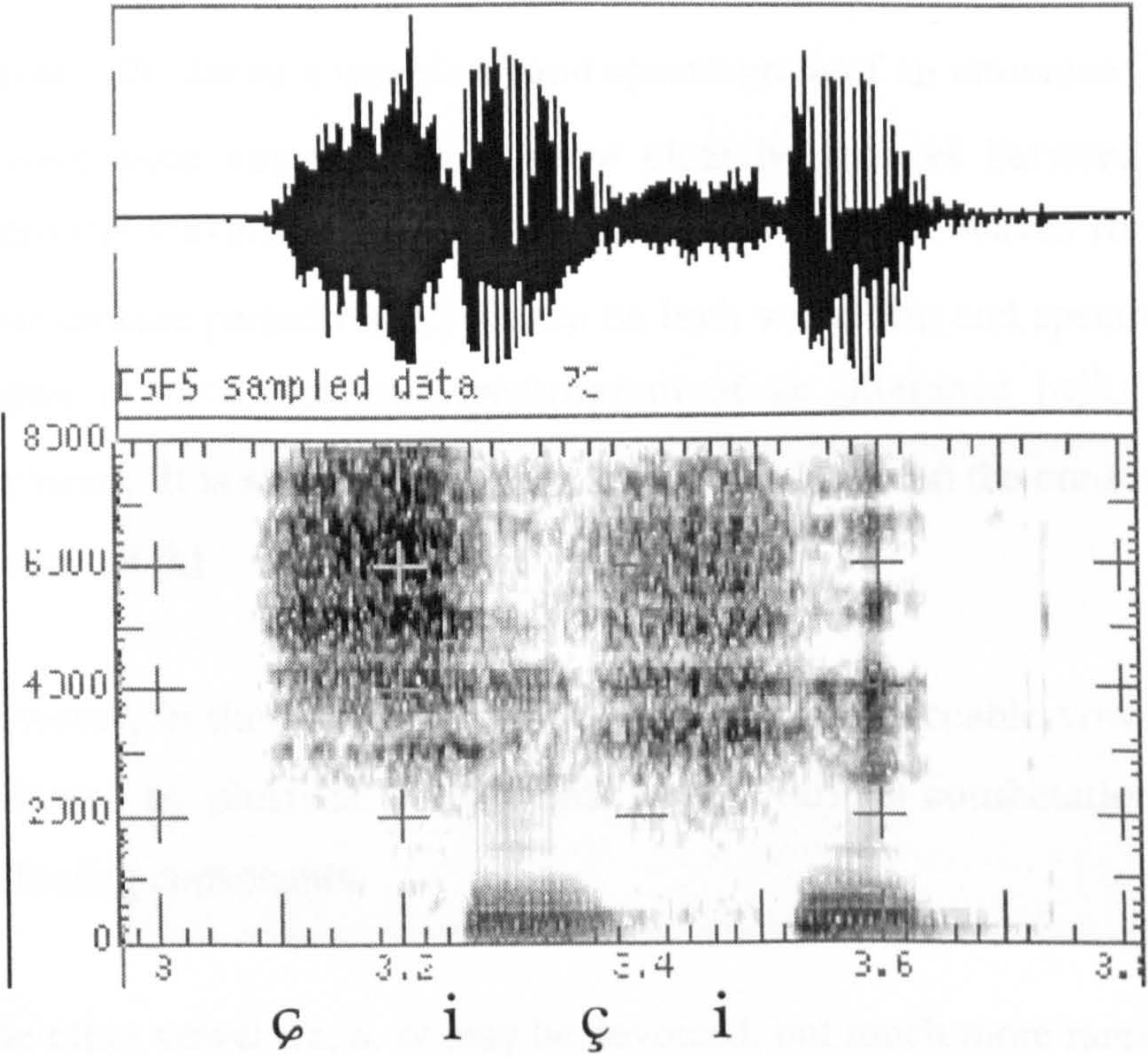
Figures 5.1a and 5.1b show spectrograms and waveforms of a word *shihi* /sihi/ (no accent) 'self financed' in two different pronunciations.

Figure 5.1a shows a pronunciation of [ɕiçi] with fully voiced vowels [i]s. Both spectrogram and waveform show clear boundaries between voiceless fricatives and voiced vowels: clear striations on the spectrogram and periodic waveform for the vowels, and aperiodic waveform for the voiceless fricatives [ɕ] and [ç]. Figure 5.1b is a spectrogram and waveform of [ɕjçj] where both vowels are voiceless. The waveform show a continuation of aperiodic waveform and the spectrogram shows no striations. As a result, the boundary between [ɕ] and [ç] is very difficult to locate although the consonants [ç] and [ɕ] have different frequencies of energy concentration: [ɕ] has concentration of higher frequency, and [ç] has a concentration of lower frequency and also has weak lower turbulences.

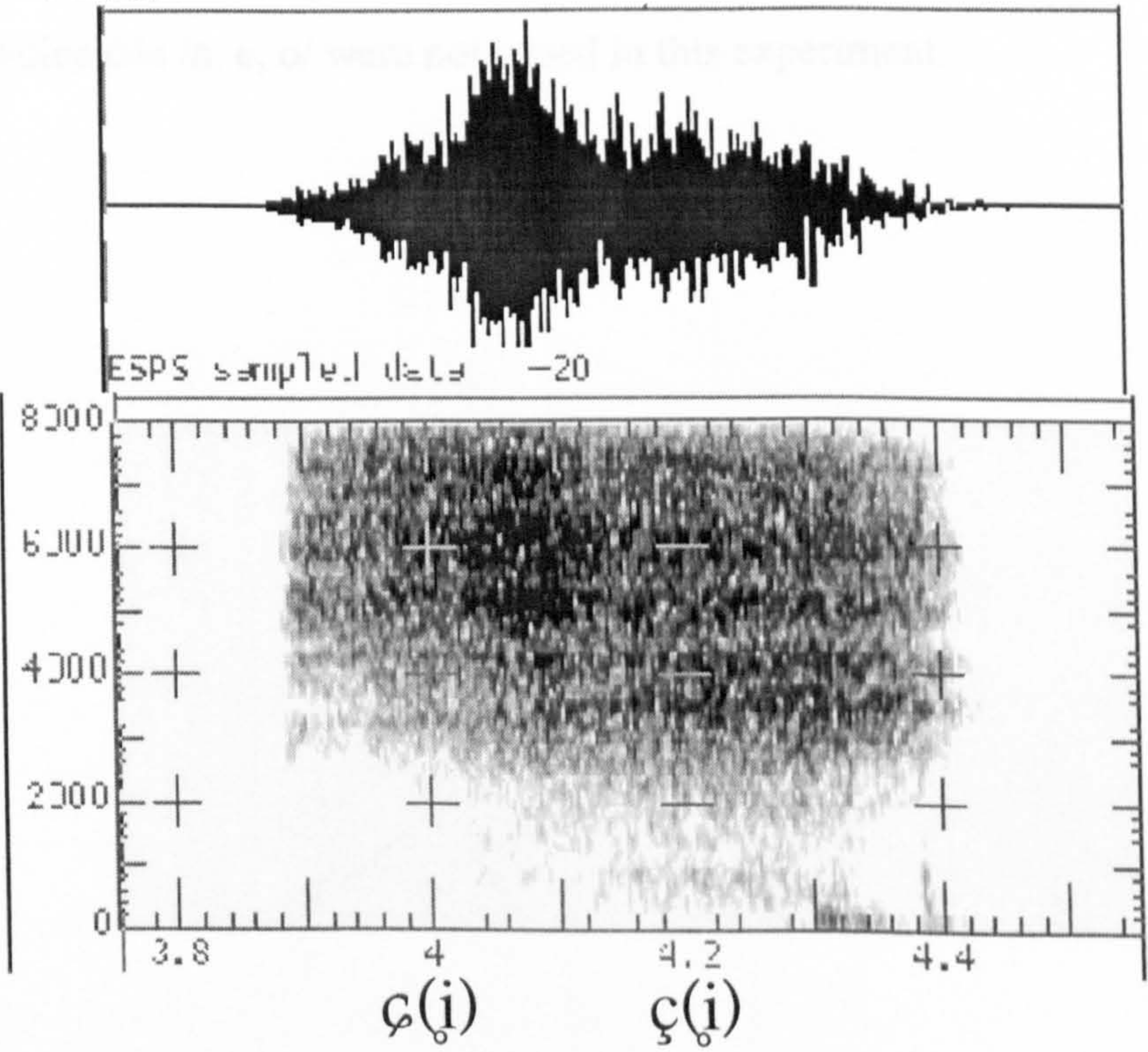


Figures 5.1a and 5.1b    Spectrograms and waveforms of two different pronunciations of a word *shihi* /sihi/ 'self financed'

(5.1a) [çiçi]



(5.1b) [ç̞iç̞i]





However, when a devoiceable vowel is followed by a voiceless plosive, on the other hand, even when the vowel is devoiced it is still easier to locate a segmental boundary between voiceless consonants as shown in Figures 5.2a and 5.2b.

Figure 5.2a shows a waveform and spectrogram of an utterance [ɕiki] both of whose vowels were voiced. They show clear boundaries between the segments; the aperiodic waveform for [ɕ] is followed by periodic waves for the first [i], then a clear closure period for [k] is seen on both waveform and spectrogram. Figure 5.2b shows a waveform and spectrogram of an utterance [ɕikɿ] with both vowels devoiced. It is still easy to locate a boundary between the consonants because of the closure of [k].

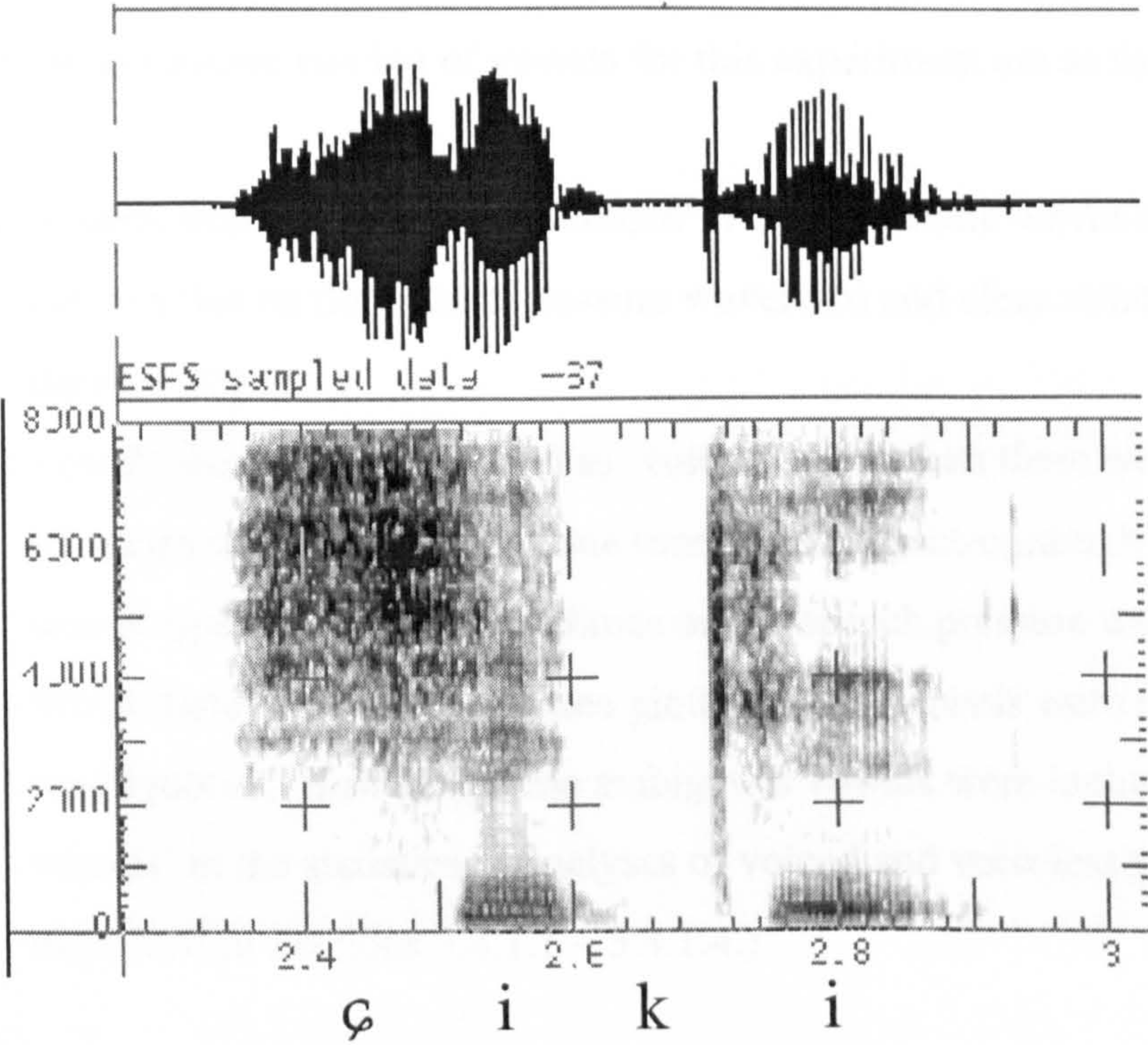
However, in the consecutive devoicing sites, devoiceable vowels were not always followed by plosives. Test words permit various combinations of preceding and following consonants.

The other vowels /e, a, o/ may be devoiced, but much more rarely, this being limited to certain phonological environments (see Chapter 4, Section 4.5.1). Therefore, devoiceable /a, e, o/ were not tested in this experiment.

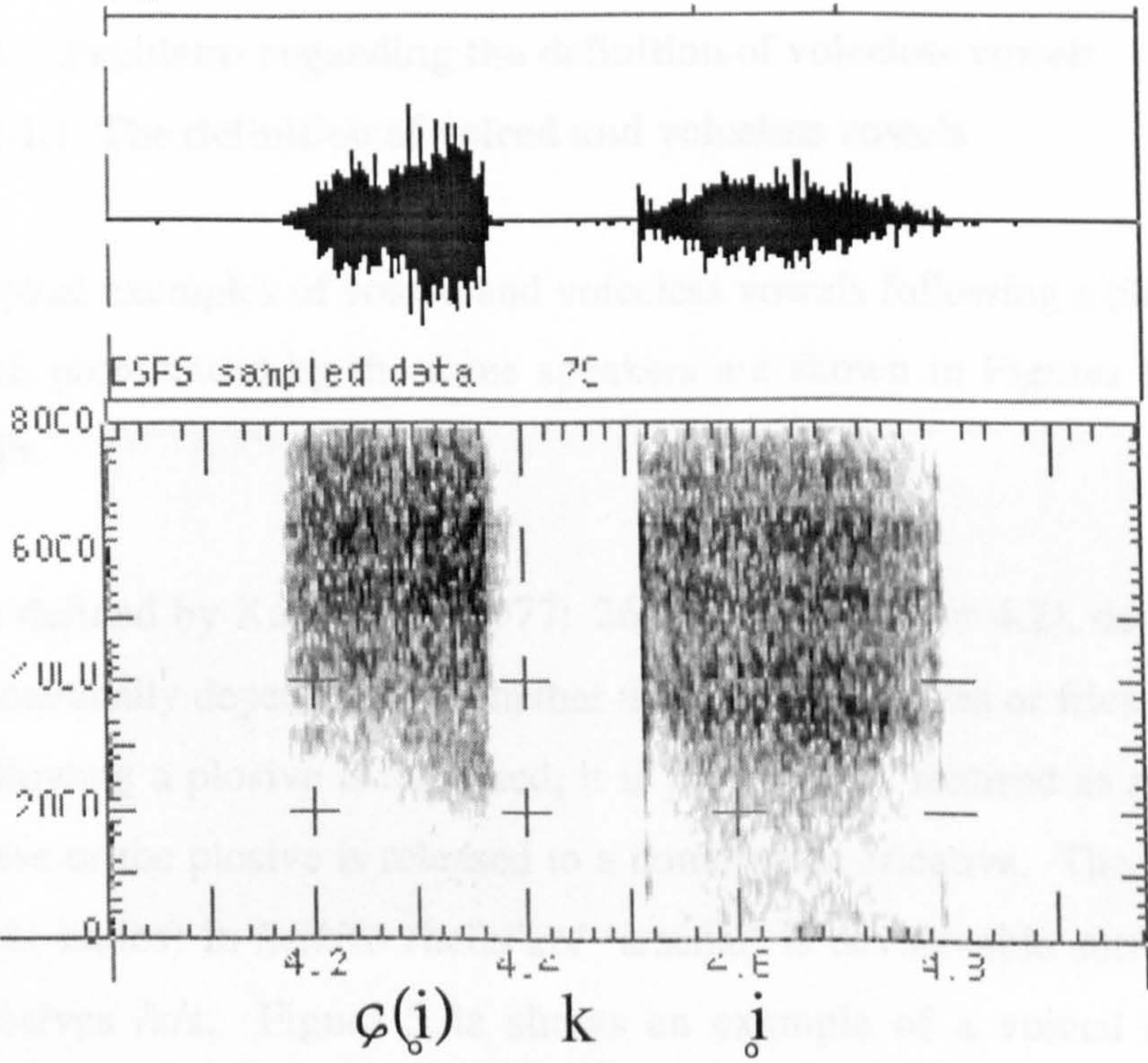


Figures 5.2a and 5.2b Spectrogram and waveform of two different pronunciations of a word *siki* /siki/ 'conduct'

(5.2a) [ɕiki]



(5.2b) [ɕikj]





### 5.3.3 The criteria for determining voicing

Vowel devoicing was examined by both waveforms and spectrograms. The criteria used to determine voicing of vowels for this experiment are as follows:

- (a) Vowels were categorised as ‘voiced’ if clear periodic waves appeared for at least three cycles on the speech pressure waveform and clear striations were visible on the spectrogram.
- (b) Vowels were still categorised as ‘voiced’ even when there were no clear striations repeating at least three times on the spectrogram, if clear periodic waves repeated at least three times on the speech pressure waveform.
- (c) When there were less than three glottal cycles, vowels were categorised as ‘ambiguous’. However, these ambiguous vowels were included in ‘voiceless vowels’ in the statistics. (Analyses of voiced and voiceless vowels will be fully discussed in Sections 5.4.1.1 ~ 5.4.1.4.)

## 5.4 Results and discussion

### 5.4.1 Problems regarding the definition of voiceless vowels

#### 5.4.1.1 The definition of voiced and voiceless vowels

Typical examples of voiced and voiceless vowels following a plosive and a fricative both pronounced by the same speakers are shown in Figures 5.3a, 5.3b, 5.4a and 5.4b.

As defined by Kawakami (1977: 26, also see Chapter 4.2), devoiced vowels differ phonetically depending on whether they follow plosives or fricatives. When a vowel following a plosive is devoiced, it is phonetically realised as a long aspiration-like noise or the plosive is released to a homorganic fricative. The second /i/ (written in bold italics) in *heikiko* /heikī<sup>h</sup>ko/ ‘arsenal’ is devoiceable surrounded by voiceless plosives /k/s. Figure 5.3a shows an example of a voiced [i] in the devoicing environment. There is a clear closure period for voiceless plosive [k] following a



long vowel [e:], then followed by a burst of [k], then an aspiration period. Immediately after the aspiration of [k], the speech pressure waveform clearly shows periodicity repeating about ten times for the [i], and there are clear striations shown on the spectrogram.

Figure 5.3b shows a typical example of voiceless [j] after a voiceless plosive in an utterance of /heiki<sup>h</sup>ko/. Similar to Figure 5.3a, a clear closure period for the voiceless plosive [k] is seen on both the speech waveform and spectrogram. However, there are no periodic waveforms and no striations on the spectrogram after the release of [k]. An aspiration-like aperiodic waveform continues until the closure of the second [k] on the waveform, and there is a concentration of noise above around 3 to 4 KHz on the spectrogram, which is a characteristic of a palatal fricative rather than a velar fricative.

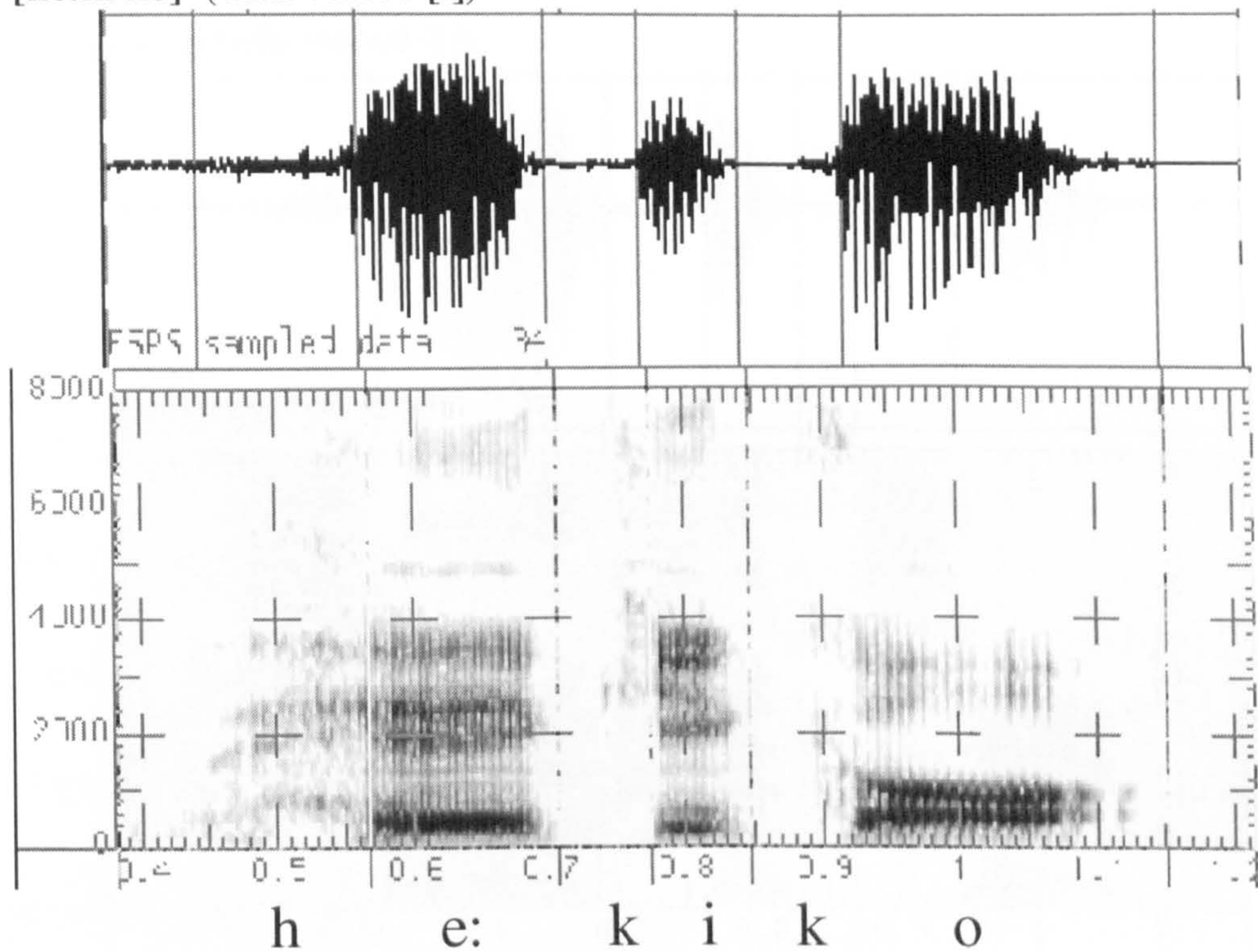
When a vowel following a fricative or affricate is devoiced, on the other hand, it becomes a continuation of a preceding fricative. Figure 5.4a shows a typical example of a voiced vowel [i] following a voiceless fricative in an utterance of /sjooh<sup>h</sup>ti/. The periodic waveform of the voiced [i] follows the aperiodic waveform for the voiceless fricative [ç] and continues until the closure of the following [t], and clear striations of the [i] are shown on the spectrogram.

Figure 5.4b shows an example of devoiced [j] after a voiceless fricative. There are no periodic waveforms or striations for the vowel /i/. Instead, the preceding fricative [ç] continues until the closure of the following [t].

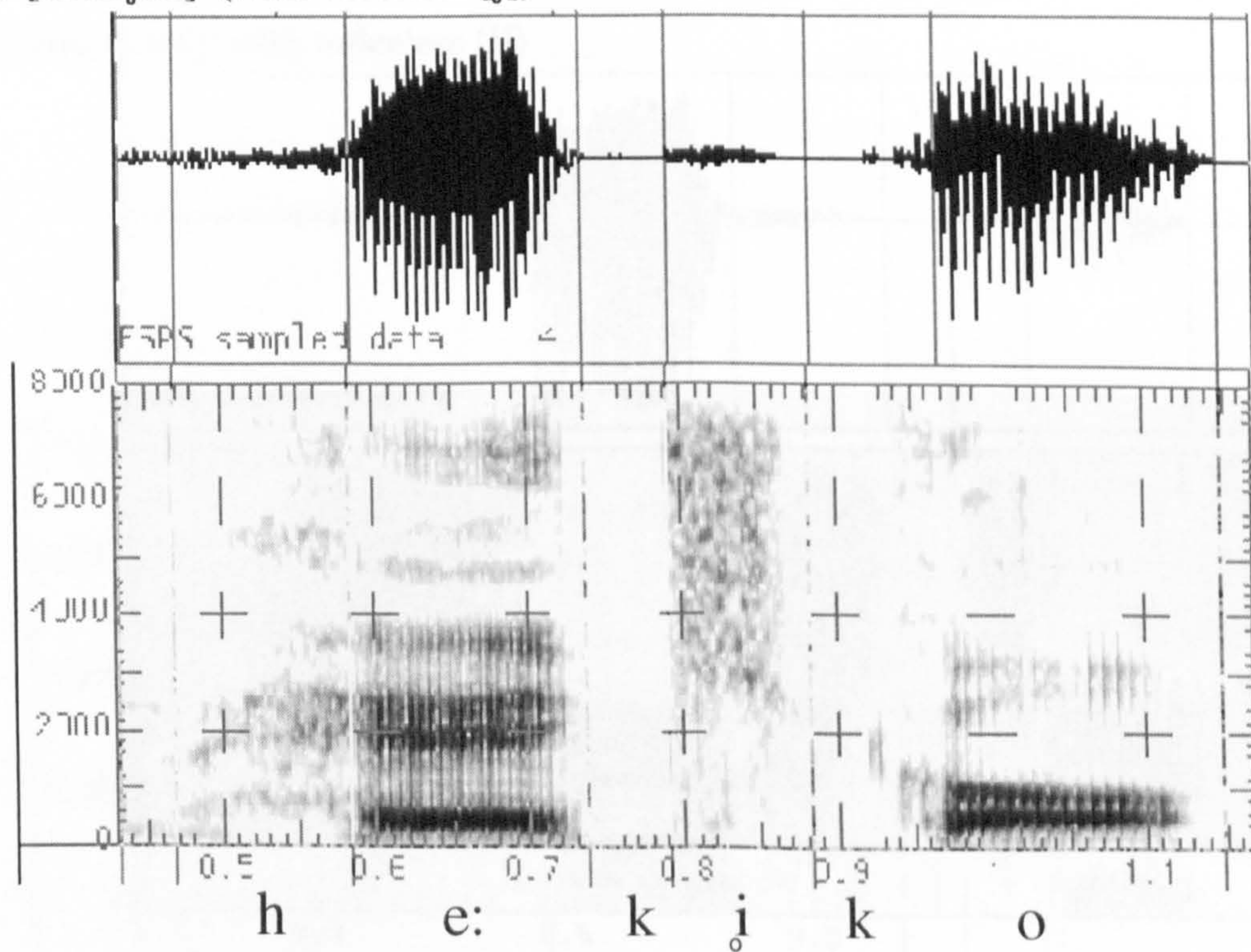


**Figures 5.3a and 5.3b** Devoiceable vowel preceded by a plosive  
*heikiko* /heiki'ko/ 'arsenal' (by Subject NM)

**(5.3a)** [he:ki'ko] (with voiced [i])



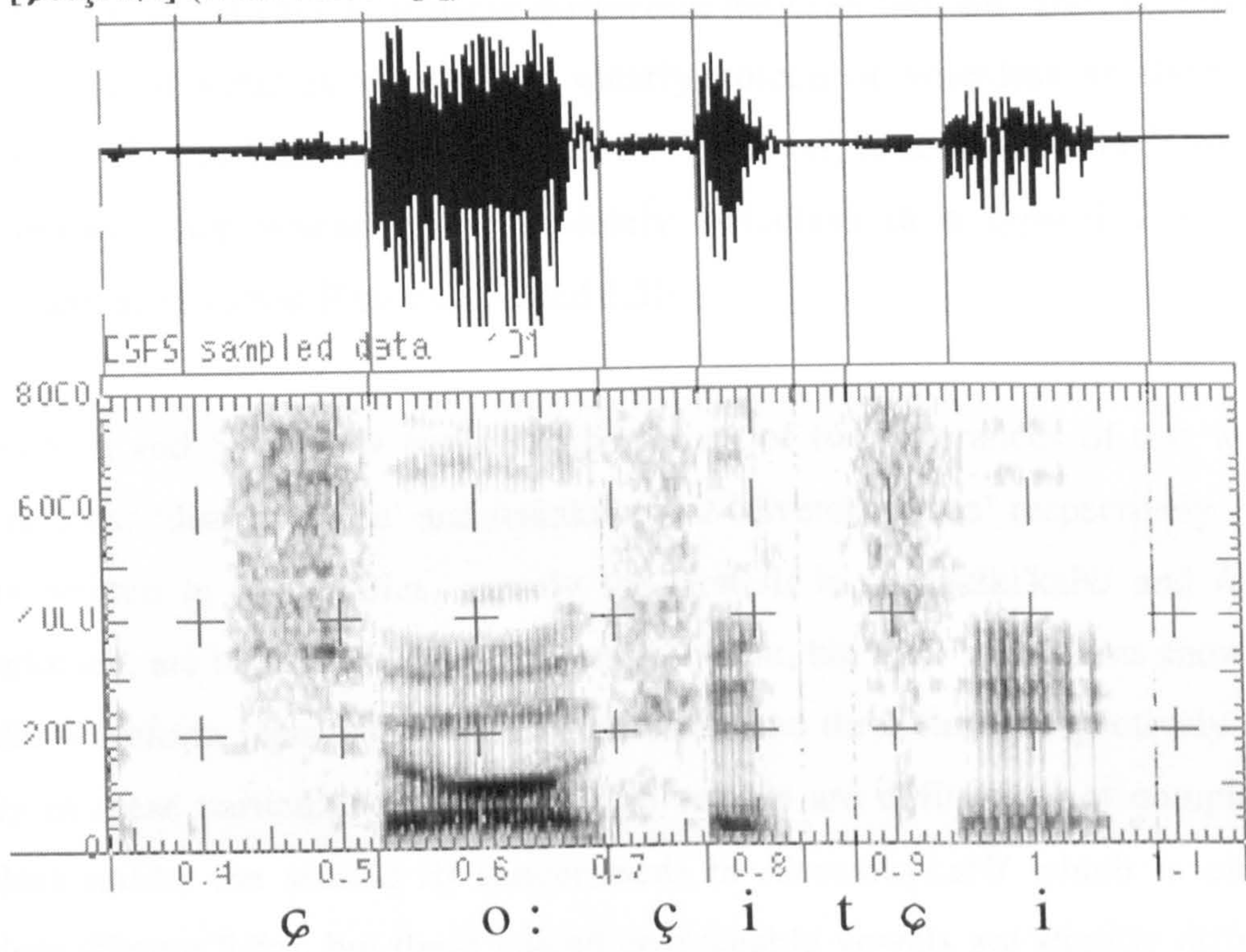
**(5.3b)** [he:ki'ko] (with voiceless [j])



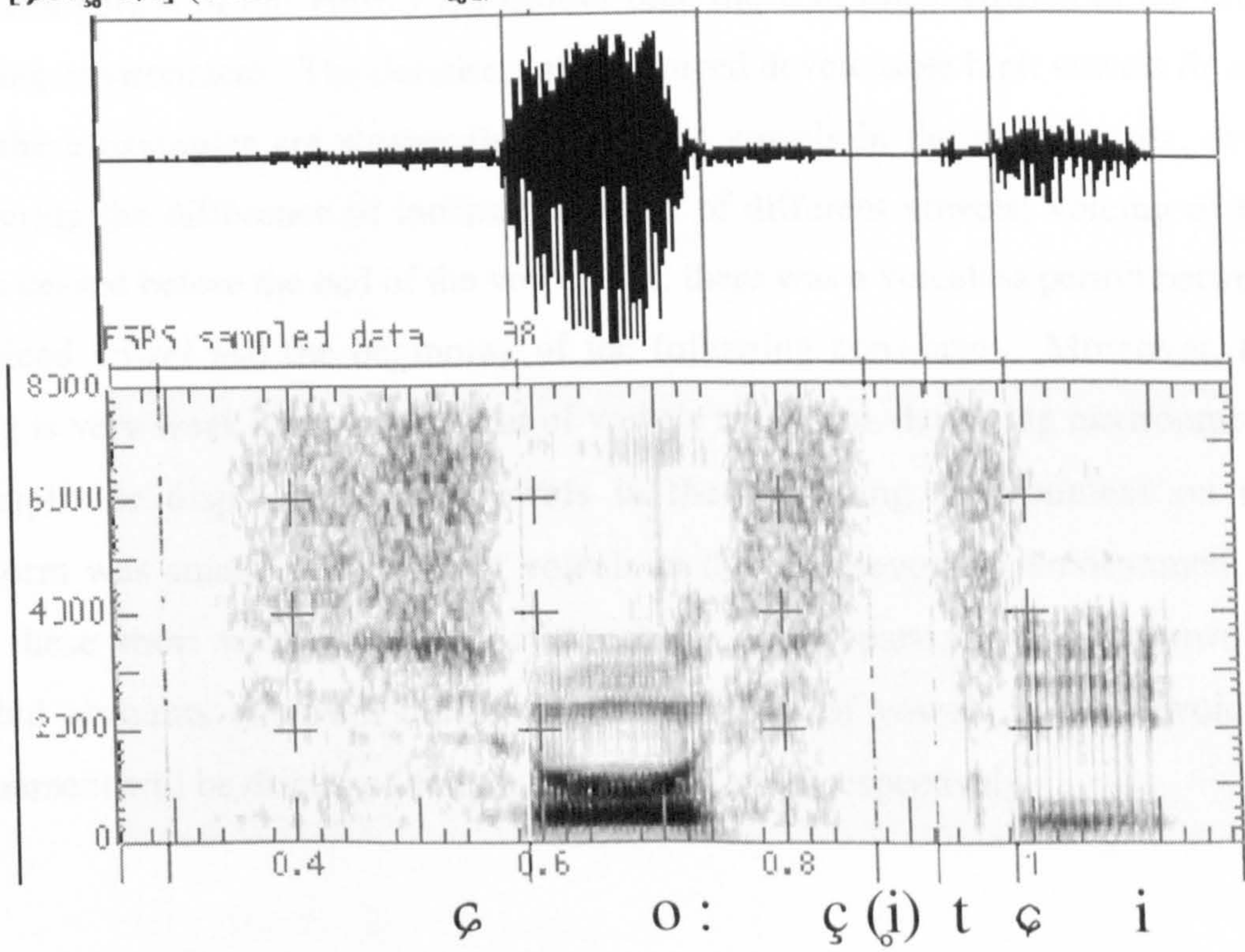


Figures 5.4a and 5.4b Devoiceable vowel preceded by a fricative  
*shoohichi* /sjoohi'ti/ 'consumer belt' (by Subject NM)

(5.4a) [ço:çi'tçi] (with voiced [i])



(5.4b) [ço:ç(i)'tçi] (with voiceless [i])





### 5.4.1.2 Partially voiced vowels

There were some problems related to the analysis of voicing. The reality of vowel devoicing was not always a clear-cut distinction between ‘voiced’ and ‘voiceless’. The majority of samples were either clearly voiced or voiceless as shown as examples in Figures 5.3a, 5.3b, 5.4a and 5.4b. However, in some cases high vowels were neither fully voiced nor completely voiceless in a typical devoicing environment as shown in Figures 5.5a and 5.5b.

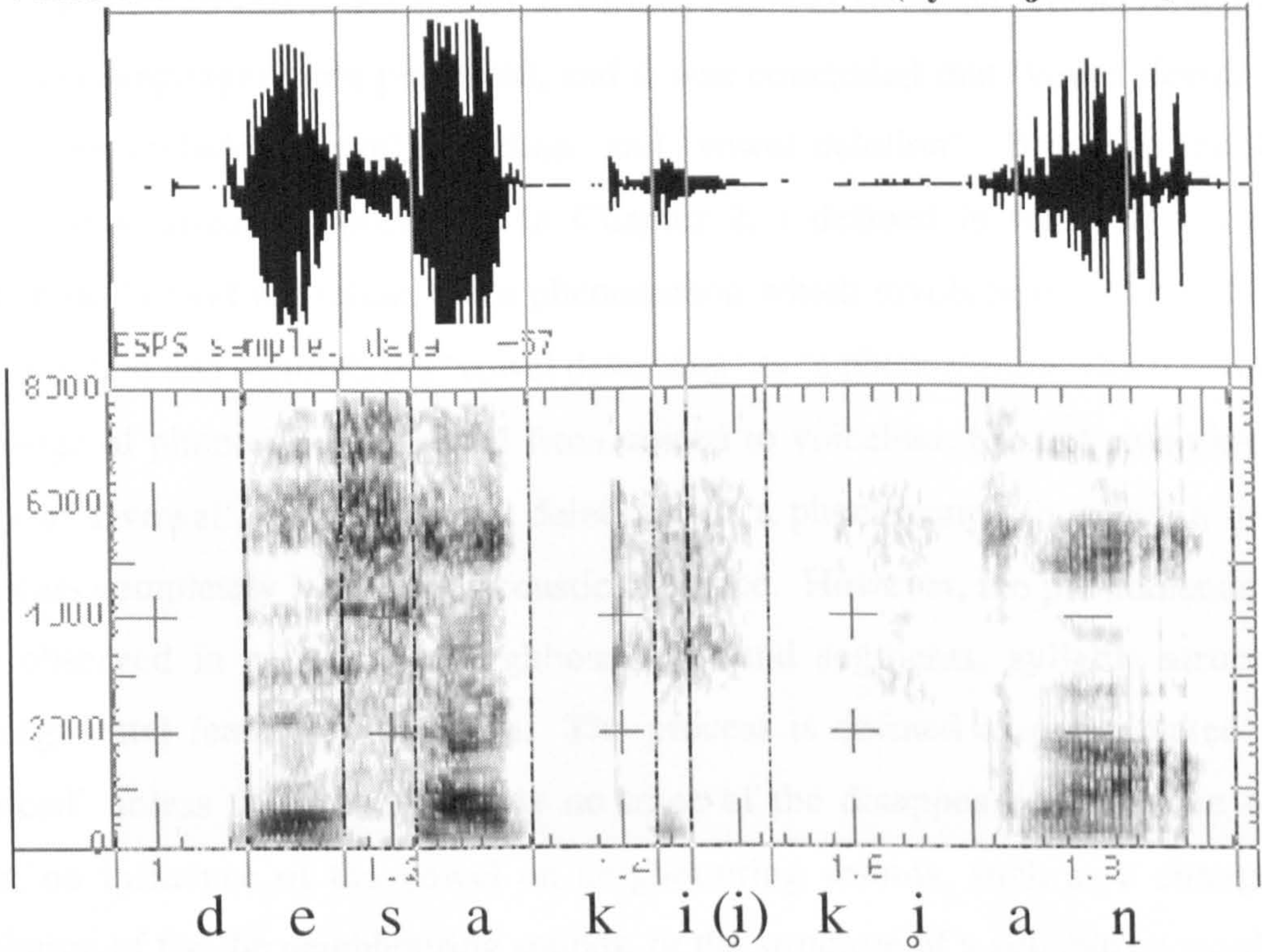
Figures 5.5a and 5.5b show periodic waveforms of two utterances of test words /desakiki`kaN/ ‘district office’ and /takakuke`iei/ ‘diversification’ respectively. The vowels written in bold italics, namely the first /i/ in /desakiki`kaN/ and /u/ in /takakuke`iei/, are both in the devoicing environment, but their waveforms showed a periodic waveform repeating around seven times and three times respectively very weakly in these particular utterances. The vowels are definitely not completely voiceless unlike the second /i/ (underlined) in /desakiki`kaN/ which is clearly voiceless (Figure 5.5a), but these voiced devoiceable vowels are slightly different from other vowels in the non-devoicing environment. First of all, the durations of the voiced period of the vowels are shorter than those of other vowels in the non-devoicing environment. The durations of the voiced devoiceable high vowels /i/ and /u/ in these examples are shorter than the other vowels in the same words, even considering the difference of intrinsic duration of different vowels; voicing of the vowels ceased before the end of the vowels, i.e. there was a voiceless period between the voiced vowel and the beginning of the following consonant. Moreover, the voicing is very weak compared to that of vowels in the non-devoicing environment; the amplitude displacement of vowels in the devoicing environment on the waveform was smaller than that of vowels in the non-devoicing environment. I name these short weakly voiced devoiceable vowels ‘partially voiced vowels’. Detailed accounts of vowel duration and amplitude of vowels in the devoicing environment will be discussed fully in Chapters 7 and 8 respectively.



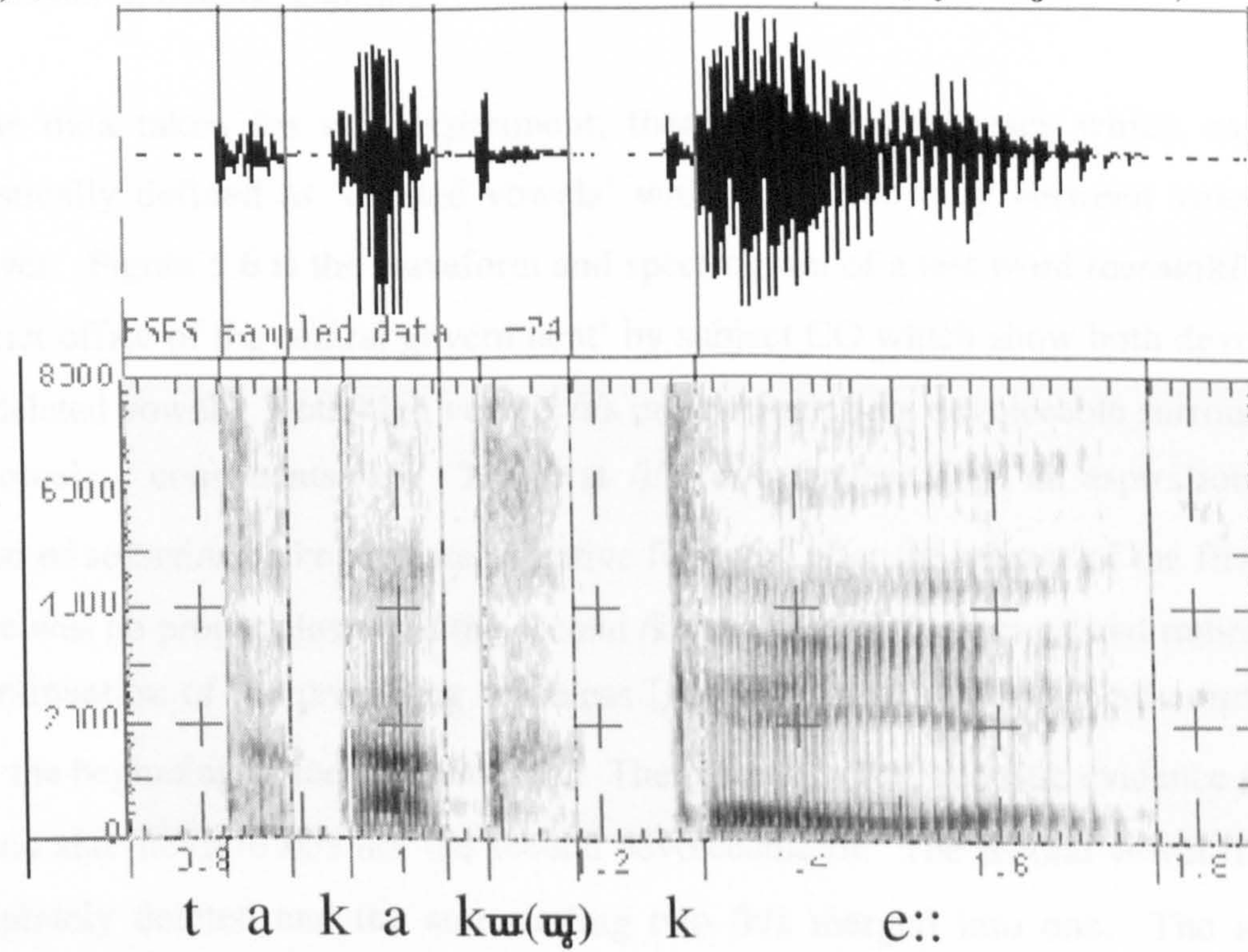
Figures 5.5a and 5.5b

Examples of speech pressure waveforms of vowels (written in *italics*)  
in the devoicing environment which are not fully voiced

(5.5a) /desakiki'kaN/ 'district office' (by Subject CO)



(5.5b) /takakuke'iei/ 'diversification' (by Subject TM)





### 5.4.1.3 Vowel deletion

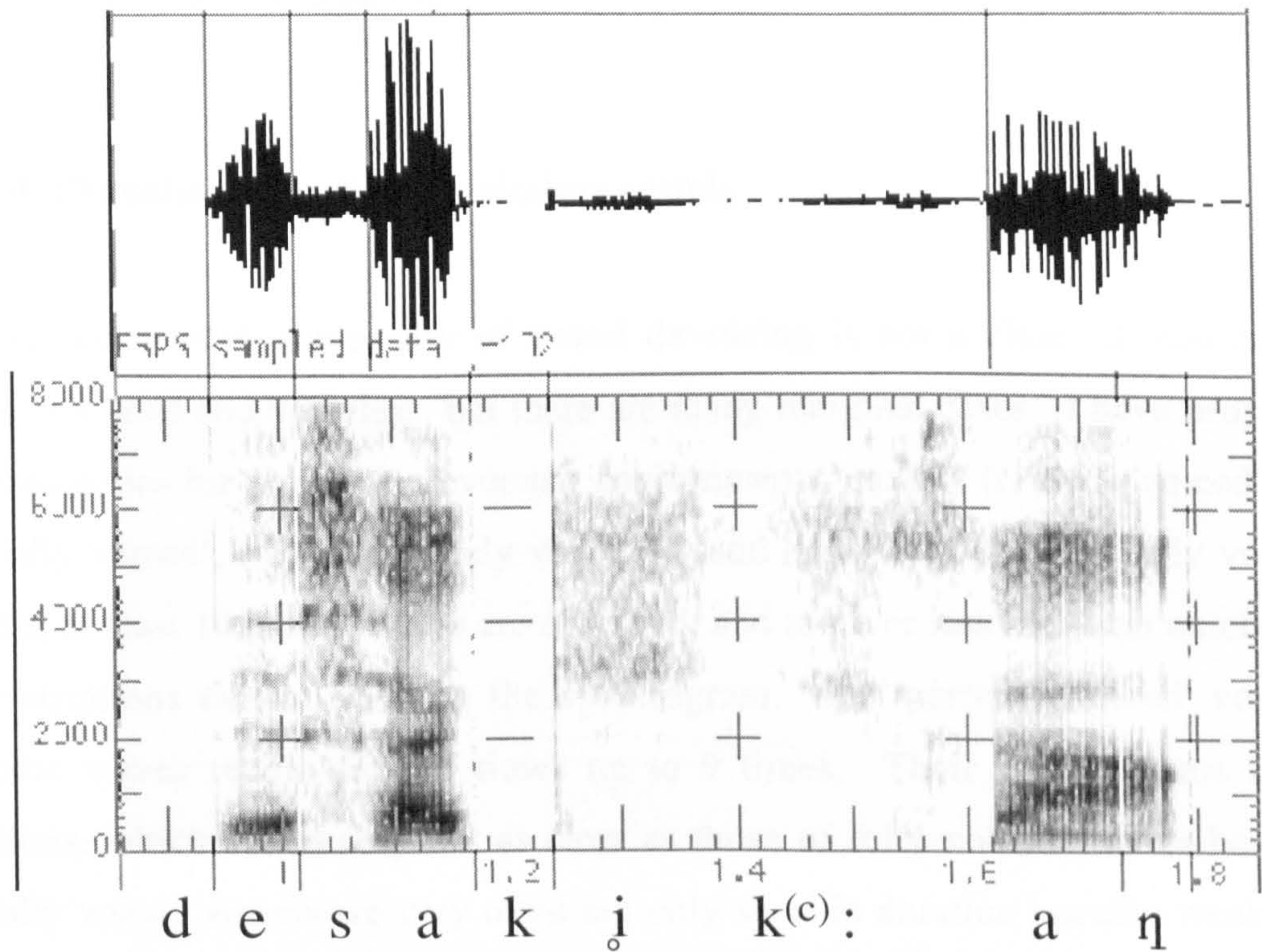
There is another category of vowels which occur in the devoicing environment, namely ‘deleted vowels’. In Chapter 2, various cases of so called ‘vowel devoicing’ in different languages were presented, and it was concluded that ‘vowel devoicing’ in fact often includes ‘vowel reduction’ and ‘vowel deletion’. Based on the data from various languages presented in Chapter 2, I defined in terms of acoustic evidence (a) ‘vowel reduction’ as ‘a phenomenon which involves reduction of both vowel duration and quality’, (b) ‘vowel devoicing’ as ‘a phenomenon which involves the change of phonation of a vowel from voiced to voiceless, but not changing the quality of a vowel’, and (c) ‘vowel deletion’ as ‘a phenomenon in which a vowel disappears completely leaving no acoustic evidence. However, the phenomenon has to be observed in relation to neighbouring sound segments, syllable structure, suprasegmental features, and so on. The process is defined as not ‘deleted’ but ‘devoiced’ unless there is absolutely no trace of the disappeared vowel, in other words, no influence of the vowel on neighbouring sounds, such as a change of articulation of the the neighbouring sounds, or the structure of a syllable or word, for example, no accent shift, no trace of the vowel in terms of duration, etc. (for details, see Chapter 2, Section 2.2.2).

In the data taken for this experiment, there were a few cases which can be acoustically defined as ‘deleted vowels’ which occurred only between voiceless plosives. Figure 5.6 is the waveform and spectrogram of a test word /desak*i*k*i*kaN/ ‘district office of the central government’ by subject CO which show both devoiced and deleted vowels. Both high vowels /i/s in bold italics are devoiceable surrounded by voiceless consonants /k/s. The first /i/ was voiceless ([j]); an aspiration-like period of something like a palatal fricative followed after the release of the first [k]. There was no proper closure of the second /k/; it was slightly fricated and rather like a continuation of the preceding voiceless [j]. This fricative-like sound continued until the beginning of the last vowel /a/. There was no clear acoustic evidence of the second and the third /k/s nor the second devoiceable /i/. The second vowel /i/ was completely deleted and the surrounding two /k/s merged into one. The actual



pronunciation was [desak̚k̚iːˈaŋ]. This type of 'vowel' may be defined as 'deleted'. There were 13 examples of deleted vowels in the experiment as listed on Table 5.3.

**Figure 5.6** An example of vowel deletion  
/desakikiˈkan/ 'district office of the central government' by subject CO



**Table 5.3** List of ‘deleted’ vowels

N.B. (i) ‘Deleted’ vowels are underlined in the test words.  
(ii) ‘Devoiced’ vowels are marked with the symbol “<sub>o</sub>” underneath.  
(iii) For gloss, refer to Tables 5.1a and 5.1b.

Test word	Phonetic realisation	Number	Subject
/desak̚k̚iːkaN/	[desak̚k̚iːˈaŋ]	2	CO
/desak̚k̚iːkaN/	[desac̚ːˈaŋ] <sup>5</sup>	1 x 2 vowels	NM
/desak̚k̚iːkaN/	[desak̚iːˈkaŋ]	1	MO
/suizoˈk̚ukaN/	[suizoˈk̚ːaŋ]	3	HM
		2	CO
/sigak̚kaˈikaN/	[sigak̚ːaˈikaŋ]	3	HM

<sup>5</sup> The first /k/ was fricated.



It must be noted that in order for this type of vowel to be defined as 'deleted', other factors such as the duration of the syllable or a whole word must be taken into account, but it is not clear at this stage whether the duration of the vowel is retained or not (durational aspect of morae with devoiced and deleted vowels will be discussed in detail in Chapter 7).

#### 5.4.1.4 Phonetic realisation of voiceless vowels

As has been stated, the reality of vowel devoicing is not a clear cut distinction between voiced and voiceless, but there are many marginal cases. I have proposed four categories for vowels in devoicing environments, namely (i) 'fully voiced', (ii) 'partially voiced', (iii) 'completely voiceless' and (iv) 'deleted'. For 'fully voiced' vowels, at least 10 glottal cycles are observed, and more or less the same number of clear striations can be seen on the spectrogram. For 'partially voiced' vowels, periodic waves repeat from 3 times up to 9 times. Their spectrograms show striations, which are usually not as clear as those of fully voiced vowels because partially voiced vowels are very often not only short in duration but also weaker in intensity. There is one problem with regard to determining vowel devoicing: i.e. where to locate a boundary between each category. It is very difficult to draw a boundary between fully voiced vowels and partially voiced vowels. The difficulty can be expressed as 'what is the difference between vowels with eleven repeated periodic waves and vowels with nine repeated periodic waves?'. The answer is 'cannot be determined for certain'. Therefore, if the waveform shows definitely a clear sign of voicing, i.e. a few periodic waves, the vowel is categorised as 'voiced'.

There is another category of voicing which was mentioned briefly in the method section (Section 5.3): namely 'ambiguous' vowels, meaning it is not clear whether even part of the vowels is voiced or not. These vowels show only one or two pulses on the waveform and usually no clear striations can be seen on the spectrogram. These pulses are extremely weak and it is not known whether they are very weak



waves or incomplete periodic waves, or just a reflection of some noise in the vocal tract. There were only 6 examples of ambiguous cases out of 1327 vowel samples analysed for this experiment. Figure 5.7 shows an example of 'an ambiguous' vowel. The second /u/ seems to be devoiced, but a couple of very weak pulses can be seen on the waveform. Although these 'ambiguous' vowels may not be completely voiceless, neither can they be categorised as 'voiced'. It is because the one or two pulses are too weak to be considered as a full periodic wave, and no striations are observed on the spectrogram. Therefore, the 'ambiguous' vowels were categorised as 'voiceless' vowels in this experiment.

**Figure 5.7** An example of an ‘ambiguous’ vowel  
**/kutu (ko’odjoo)/ ‘shoe (factory)’ (by Subject NM)**

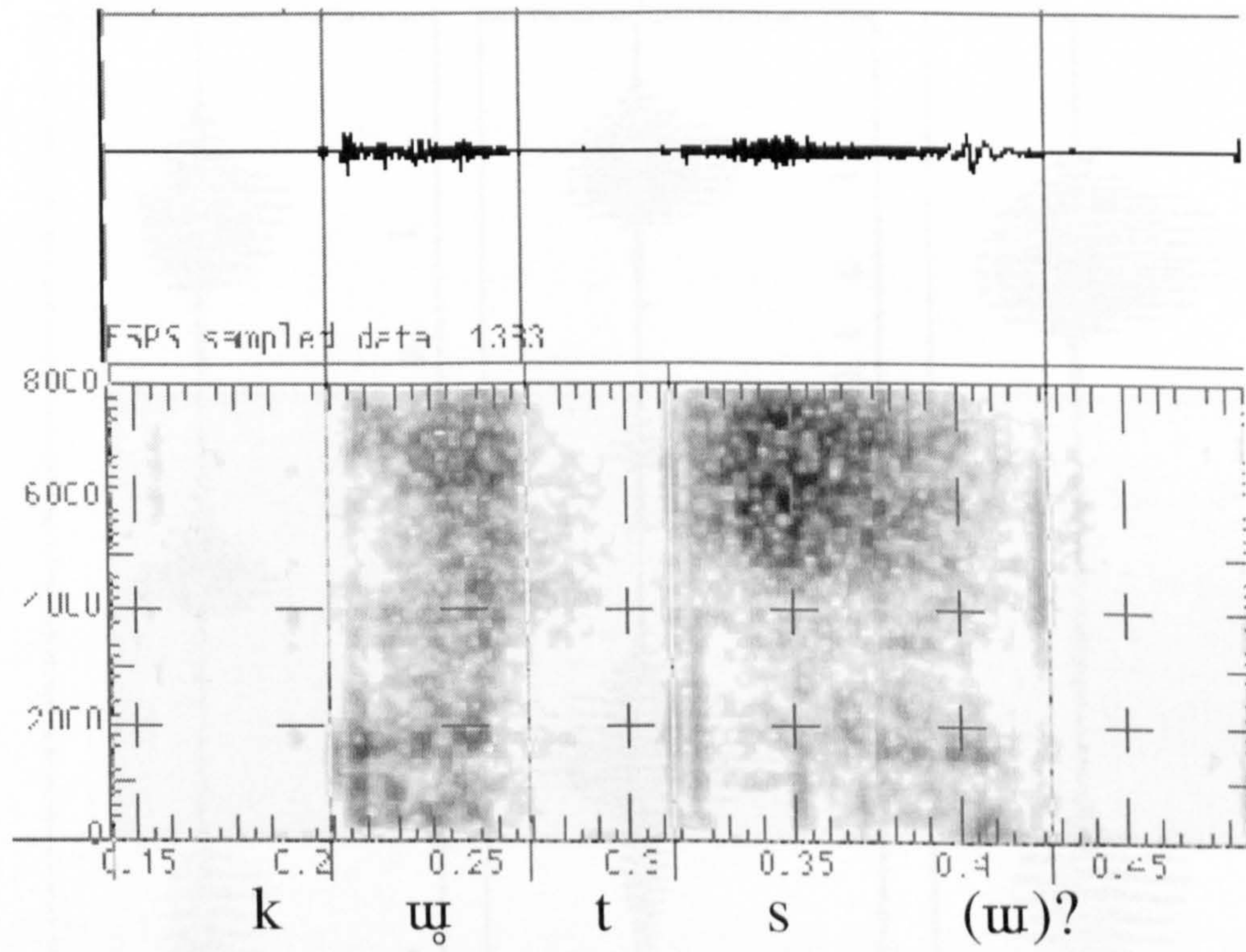
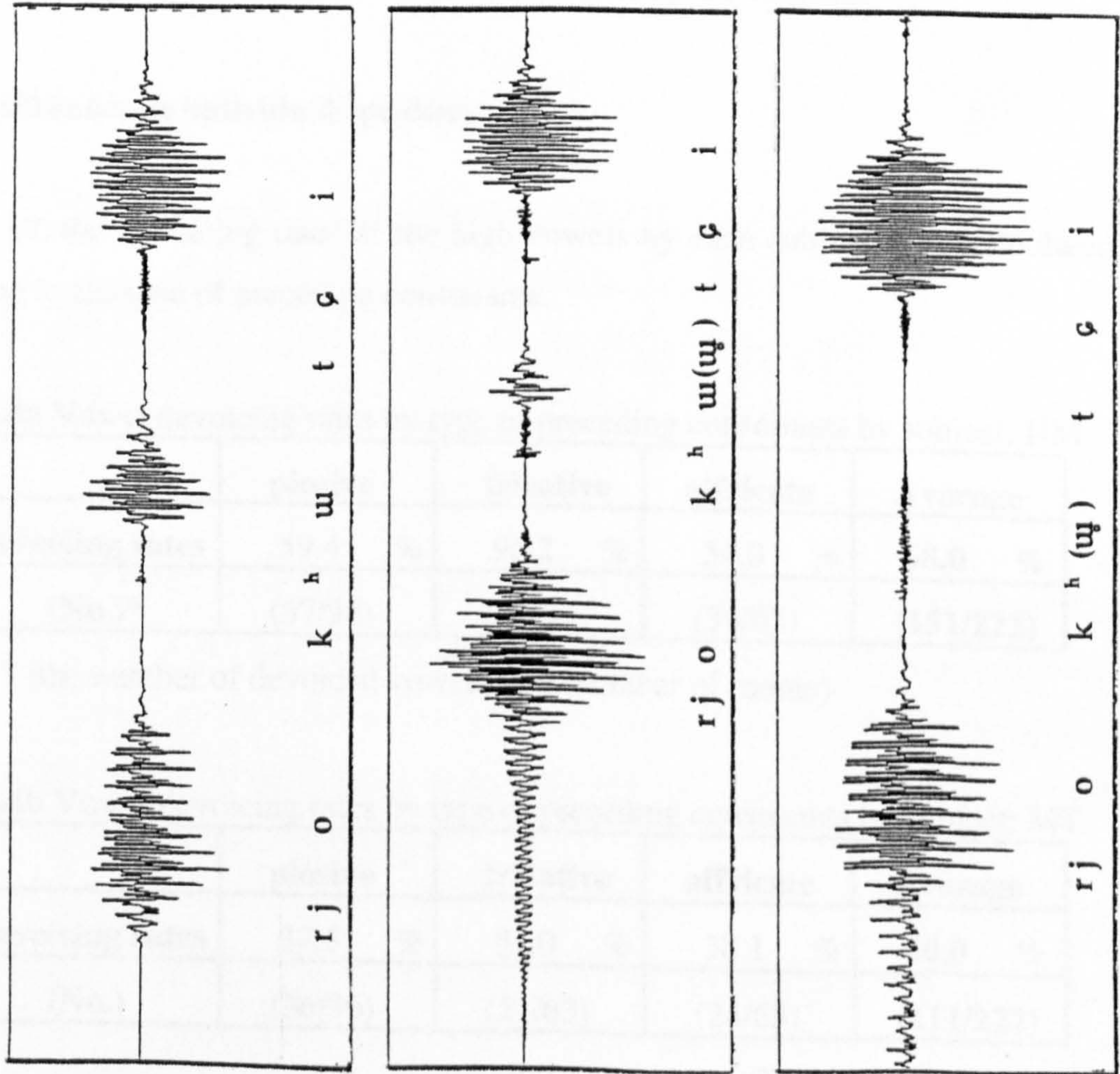




Figure 5.8 Phonetic realisation of devoiceable vowels

/rjokuti(kooeN)/ 'forest(park)'



fully voiced /u/

partially voiced /u/

completely voiceless /u/



Vowel devoicing in reality does not involve a categorical distinction between voiced and voiceless, but a gradual change of phonation from fully voiced to completely voiceless, or perhaps the vowel may be deleted completely. As shown in Figure 5.8, the realisation of devoiceable vowels can be anything between the two extreme categories.

Vowel duration will be discussed in Chapter 7 and the intensity of the vowel will be discussed in Chapter 8 in detail.

5.4.1.5 Accent placement

There were some cases in which the subjects put an accent on vowels that was different from the ones listed in The NHK Accent Dictionary (1985). In those cases, the analyses were based on the actual pronunciations by the subjects.

5.4.2 Difference in individual speakers

First of all, the devoicing rates of the high vowels by each subject were calculated according to the type of preceding consonants.

Table 5.4a Vowel devoicing rates by type of preceding consonants by subject: HM

	plosive	fricative	affricate	Average
devoicing rates	59.4 %	95.2 %	54.0 %	68.0 %
(No.)*	(57/96)	(60/63)	(34/63)	(151/222)

\* (the number of devoiced vowels/total number of morae)

Table 5.4b Vowel devoicing rates by type of preceding consonants by subject: MT

	plosive	fricative	affricate	Average
devoicing rates	37.5 %	81.0 %	38.1 %	50.0 %
(No.)	(36/96)	(51/63)	(24/63)	(111/222)

(Tables 5.4 continues)

(Table 5.4 continued)

**Table 5.4c** Vowel devoicing rates by type of preceding consonants by subject: CO

	plosive	fricative	affricate	Average
devoicing rates	46.8 %	82.3 %	42.9 %	55.7 %
(No.)	(44/94)	(51/62)	(27/63)	(122/219)

**Table 5.4d** Vowel devoicing rates by type of preceding consonants by subject: TM

	plosive	fricative	affricate	Average
devoicing rates	42.7 %	61.9 %	39.7 %	47.3 %
(No.)	(41/96)	(39/63)	(25/63)	(105/222)

**Table 5.4e** Vowel devoicing rates by type of preceding consonants by subject: NM

	plosive	fricative	affricate	Average
devoicing rates	32.3 %	68.3 %	22.2 %	42.8 %
(No.)	(31/96)	(50/63)	(14/63)	(95/222)

**Table 5.4f** Vowel devoicing rates by type of preceding consonants by subject: MO

	plosive	fricative	affricate	Average
devoicing rates	52.1 %	90.5 %	54.0 %	60.8 %
(No.)	(50/96)	(57/63)	(28/63)	(135/222)

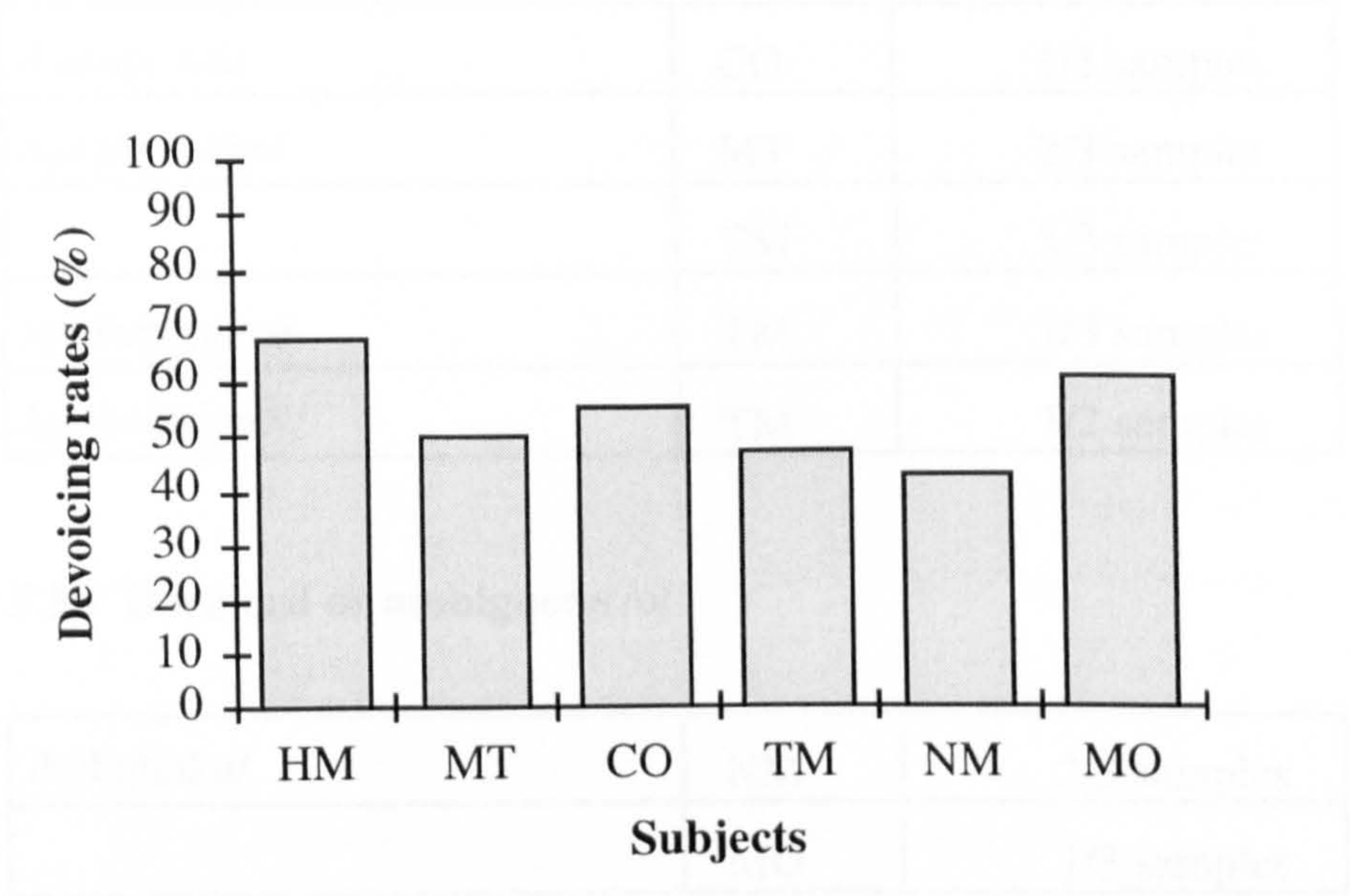
As shown in Tables 5.4a ~ 5.4f and Figure 5.9, for all subjects, the devoicing rates were the highest when voiceable vowels were preceded by fricatives than by plosives or affricates. The average devoicing rate of each subject ranged from 43% (NM) to 68% (HM), which does not seem to support that vowel devoicing is a regular process in Standard Japanese.

The difference in average devoicing rates between subjects was examined using unrelated one-way analysis of variance (ANOVA). The result suggested that there was no significant difference in devoicing rates between subjects [F(5,60), n.s.].



Therefore, individual devoicing rates were added up to obtain the average devoicing rates of six subjects, and analysed by the type of preceding consonants which is one of the controlling factors for the experiment.

**Figure 5.9 Average devoicing rates by subjects<sup>6</sup>**



**5.4.3 Non-high vowel devoicing and high vowel devoicing in non-devoicing environments**

There were a few instances of devoiced or ambiguous cases of non-high vowels /a/ and /o/, and also /i/ before voiced consonants, listed in Tables 5.5a, 5.5b and 5.5c.

<sup>6</sup> The devoicing rates include ‘ambiguous’ and ‘deleted’ vowels, but exclude ‘partially voiced’ vowels as explained in section 4.4.1.4.



**Table 5.5a Devoiced or ambiguous /a/**

test word	subject	No. of occurrence
/sjoohi-keika <u>ku</u> /	HM	2/3 samples
	CO	1/3 samples
/ka <u>s</u> upi-kai/	CO	1/3 samples
/ta <u>k</u> aku-keiei/	MT	2/3 samples
	TM	3/3 samples
/a <u>k</u> atuti-jama/	TM	2/3 samples
/a <u>s</u> ahi-kaigaN/	TM	1/2 samples

**Table 5.5b Devoiced or ambiguous /o/**

/ko <u>k</u> uhuku/	NM	2/3 samples
	MO	1/3 samples

**Table 5.5c Devoiced /i/ before voiced consonant**

/sjaapu-pens <u>i</u> ru/	CO	3/3 samples <sup>7</sup>
	MO	2/3 samples
/k <u>j</u> rihuki-ki/	TM	1/3 samples
/s <u>i</u> gaku-kaikaN/	TM	3/3 samples

**5.4.4 Adjacent consonants**

Table 5.6 shows vowel devoicing rates reanalysed by the type of preceding consonants. With regard to type of preceding consonants, devoicing rates were higher for vowels preceded by fricatives than vowels preceded by plosives and affricates. A statistical analysis by a one-way ANOVA showed that there was a

<sup>7</sup> One subject CO devoiced /r/ and /u/ at the end of the word as well.

significant difference in devoicing rates according to the types of consonants preceding devoiceable vowels [ $F(2, 8) = 6.929, p < .025$ ].

**Table 5.6 The average devoicing rates by the type of preceding consonants**

	plosive	fricative	affricate	Average
devoicing rates	45.1 %	81.6 %	40.2 %	54.0 %
(No.)*	(259/574)	(306/375)	(152/378)	(717/1327)

\*(the number of devoiced vowels/total number of samples)

There were three studies which, using a large Japanese language database, calculated vowel devoicing rates according to various phonetic environments. One was compiled by Takeda and Kuwabara (1987) and is shown in Table 5.7.

**Table 5.7 Vowel devoicing rates by preceding consonants**

(Takeda & Kuwabara, 1987: 105)

	/t/ <sup>8</sup>	/k/	/s/ <sup>9</sup>	/sj/ <sup>10</sup>	/h/ <sup>11</sup>
/i/	14 %	13 %	23 %	N/A	41 %
No.	(28/200)	(54/429)	(116/499)	N/A	(56/138)
/u/	18 %	10 %	11 %	21 %	29 %
No.	(90/499)	(77/774)	(41/366)	(27/127)	(42/156)

Their results showed lower devoicing rates for all consonant-vowel combinations than my results. This appeared to be due to the fact that their calculated figures included non-typical devoicing environments, i.e. all the above devoiceable morae were calculated even when followed by voiced sounds. They then analysed the

<sup>8</sup> /t/ is phonetically realised as [tɕ] before /i/ and [ts] before /u/.

<sup>9</sup> /s/ is phonetically realised as [ɕ] before /i/.

<sup>10</sup> /sj/ is phonetically realised as [ɕ].

<sup>11</sup> /h/ is phonetically realised as [ç] before /i/ and [ɸ] before /u/.

devoicing rates of /si/ [çi] which scored the highest number of devoiced samples according to their phonetic environment as shown in Tables 5.8 and 5.9.

**Table 5.8 The devoicing rates of /si/** (Takeda & Kuwabara, 1987: 105)

	position in a word			position of accent nucleus		
	initial	medial	final	itself	preceding vowel	others
devoicing rate	31 %	24 %	1 %	0 %	34 %	28 %
number of samples	(70/224)	(45/188)	(1/82)	(0/121)	(53/155)	(63/223)

**Table 5.9 The devoicing rate of /si/ by following consonants** (Takeda & Kuwabara, 1987: 105)

	/k/	/t/	/g/	/s/	/n/	/m/
devoicing rate	100 %	94 %	50 %	25 %	15 %	3 %
number	48	56	5	3	2	1

The important point here is that the devoicing rate of /si/ in word-final position is extremely low compared to that in word-initial and medial positions. It should also be noted that the vowel in /si/ is not devoiced when it is accented. As for the following consonant, it is not necessary for the following consonant to be voiceless. In fact, /si/ followed by /g/ was devoiced more often than followed by /s/. This may result from the tendency to avoid devoicing in /s + i + s/ sequences. This would create a long [s:] which may not be easy to perceive.

The results of the other study by Yoshida and Sagisaka (1990) are shown in Tables 5.10 and 5.11.



**Table 5.10    The rates and the number of devoiced vowels by type of preceding consonants (in /ÇVÇ/ environment)**

(Yoshida and Sagisaka, 1990: 6)

preceding consonant \ subject	a	b	c
plosive (No.)	66 % (27)	88 % (168)	81 % (156)
affricate (No.)	83 % (105)	98 % (114)	91 % (116)
fricative (No.)	77 % (256)	90 % (307)	89 % (304)

**Table 5.11    The rates and the number of devoiced /si/ by type of following consonants (in /ÇVÇ/ environment)**

(Yoshida and Sagisaka, 1990: 6)

following consonant \ subject	a	b	c
plosive (No.)	86 % (342)	99 % (392)	99 % (392)
affricate (No.)	73 % (64)	91 % (80)	93 % (82)
fricative (No.)	51 % (91)	60 % (104)	54 % (95)

Yoshida and Sagisaka's results show much higher devoicing rates than the results of Takeda and Kuwabara shown in Table 5.7. This is probably because Yoshida and Sagisaka analysed data only from typical devoicing environments. These results also show that high vowels are more likely to be devoiced when preceded by fricatives than by plosives. It should be noted that the devoicing rates for vowels preceded by affricates are higher than vowels preceded by fricatives. Considering the structure of

affricates, namely a homorganic cluster of a plosive followed by a fricative, a vowel preceded by an affricate means phonetically the same as preceded by a fricative. As for the following consonants, vowel devoicing was more common when the vowels were followed by plosives and affricates than when followed by fricatives; a vowel followed by an affricate means phonetically the same as followed by a plosive.

Nagano-Madsen (1994) studied the effect of preceding and following consonants on vowel devoicing in Japanese, using the ATR Sentence Database which consists of 503 phonemically balanced short sentences read by 10 professional announcers (6 male and 4 female). She analysed segmental devoicing environments from the tokens of one of the speakers (728 devoiced vowels), and found that high vowel devoicing rates were typically high (87.5 %) in the /ÇVÇ/ environment, but not very common (7 %) in /ÇV#/ , and also found that devoicing also occurred even when followed by voiced consonants although the rates was very low (5.8 %). She reported only one instance of vowel devoicing between voiced consonants in a sequence of /hazuda/ [hadzuɰda] '(it) must be---'. Her results from the 10 speakers showed higher devoicing rates for sequences where high vowels were preceded by fricatives and affricates, such as [çit], [çit], [tsɯk], [sɯ#] and [ɸɯk], all of which showed 100 % devoicing rates. As for the type of following consonants, it was predominantly plosives that showed high devoicing rates, but devoicing rates were lower when the following consonants were allophones of /h/, namely [h] and [ɸ]<sup>12</sup>. Nagano-Madsen considered the reason for this was that /h/ may be voiced in intervocalic position, which works as a negative effect for devoicing of the preceding vowels, as discussed in Chapter 4, Section 4.3. Other allophones of /h/, typically [ɸ] and also probably [ç], may not always articulated distinctively, and often merge with [h] and consequently accompanied with some vocal fold vibration.

The results of the above three studies explain why the devoicing rates of vowels preceded by fricatives were higher than those preceded by plosives or affricates in

---

<sup>12</sup> Nagano-Madsen stated that there was only one instance of a sequence of a devoiced vowel followed by [ç], another typical allophone of /h/. The vowel [ɯ] in the sequence [kɯç] was devoiced by all 10 speakers. However, because of the small number of tokens in the sample, following [ç] was not included in her discussion.

my experiment in which most of the vowels were followed by plosives or affricates. However, when vowels preceded by fricatives are also followed by fricatives, devoicing becomes less common than when the vowels followed by plosives or affricates.

5.4.5 Position in the utterance

Takeda and Kuwabara's results (1987) showed an extremely low devoicing rate in word final position. In my experiments, there have been examples where a high vowel was voiceless in a devoicing environment, but was voiced when a speaker paused after the vowel in question. For instance, the subject CO devoiced the second /i/ in /hitati-kaigaN/ in two out of three utterances. However, she inserted a pause at an internal word boundary between /ti/ and /ka/ in one of her utterances, and in that case /i/ in /ti/ was voiced. Also there were no instances when the vowel was usually voiced in devoicing environment and was devoiced when a pause was inserted just after the vowel. The devoicing rates obtained from my experiment shown in Table 5.6 include devoiceable vowels in word-final position. Therefore, I eliminated samples in word-final positions and recalculated devoicing rates. Slightly different results were obtained and are shown in Table 5.12.

**Table 5.12** Average devoicing rates by type of preceding consonant excluding samples in word-final positions

	plosive	fricative	affricate	Average
devoicing rates	59.1 %	82.3 %	49.7 %	64.3 %
(No.)*	(254/430)	(306/372)	(150/302)	(710/1104)

\*(the number of devoiced vowels/total number of examples)

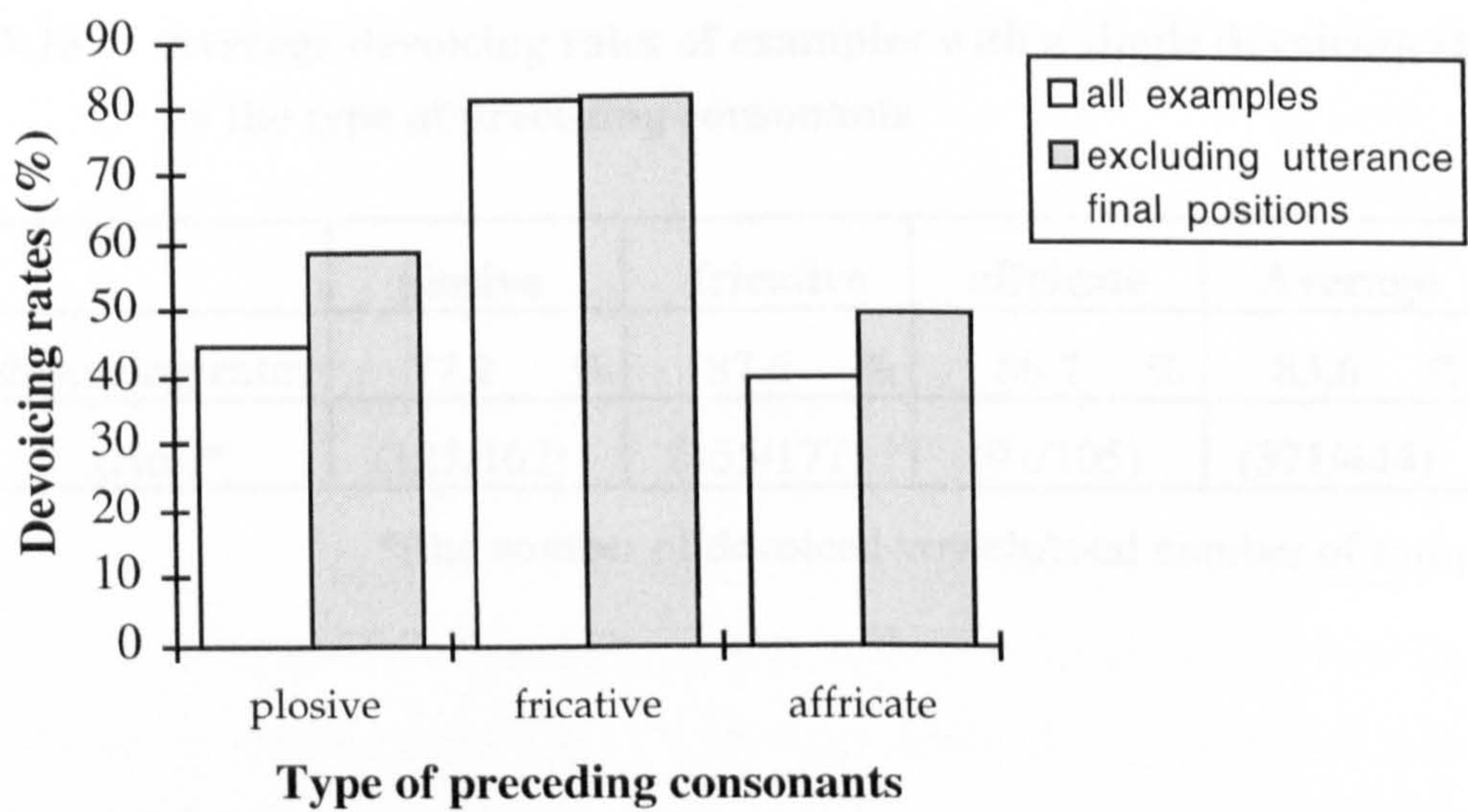
Average devoicing rates excluding examples in word final positions were compared with average devoicing rates of all examples, and are shown in Figure 5.11. Average



devoicing rates excluding examples in word final positions (Table 5.12) were slightly higher than average devoicing rates of all examples (Table 5.6).

Statistical analysis by two-way ANOVA found that the effect of vowel position in an utterance was not significant [ $F(1,16) = .396$ , n.s], but the difference between three types of preceding consonants was significant [ $F(2,16) = 10.689$ ,  $p < .025$ ]. The interaction between vowel position and type of preceding consonants was found to be not significant [ $F(2,16) = .101$ , n.s.].

**Figure 5.10** Comparison of average devoicing rates of all examples and average devoicing rates excluding examples in word final positions by the type of preceding consonant



**5.4.6 Devoicing conditions**

The third controlling factor is devoicing condition, namely whether a devoiceable vowel is in (a) the single devoicing environment, or (b) the consecutive devoicing environment. As mentioned earlier in Chapter 4, Section 4.5.8 and Section 5.2 in this chapter, in the environment where more than one devoicing site occurs consecutively, one of the devoiceable vowels (or in theory more than one when there are more than three devoicing sites) should remain voiced. The devoicing rates shown in Tables 5.6 and 5.12 include the examples in the consecutive devoicing



environment, and therefore they are not a true reflection of the analysis of devoicing. Devoicing rates must be calculated separately for words with one devoicing site and for words with more than one devoicing site. Therefore, I will calculate devoicing rates separately for words where there are only single devoicing sites and for words where consecutive devoicing is possible. Table 5.13 shows the average vowel devoicing rates of examples with a single devoicing site in non-utterance final positions, i.e. excluding examples with consecutive devoicing environments and examples in utterance final positions. Figure 5.11 shows the comparison between three different devoicing conditions, namely the average devoicing rates of (a) all examples, (b) samples excluding utterance final positions, and (c) examples excluding both utterance final positions and consecutive devoicing environments.

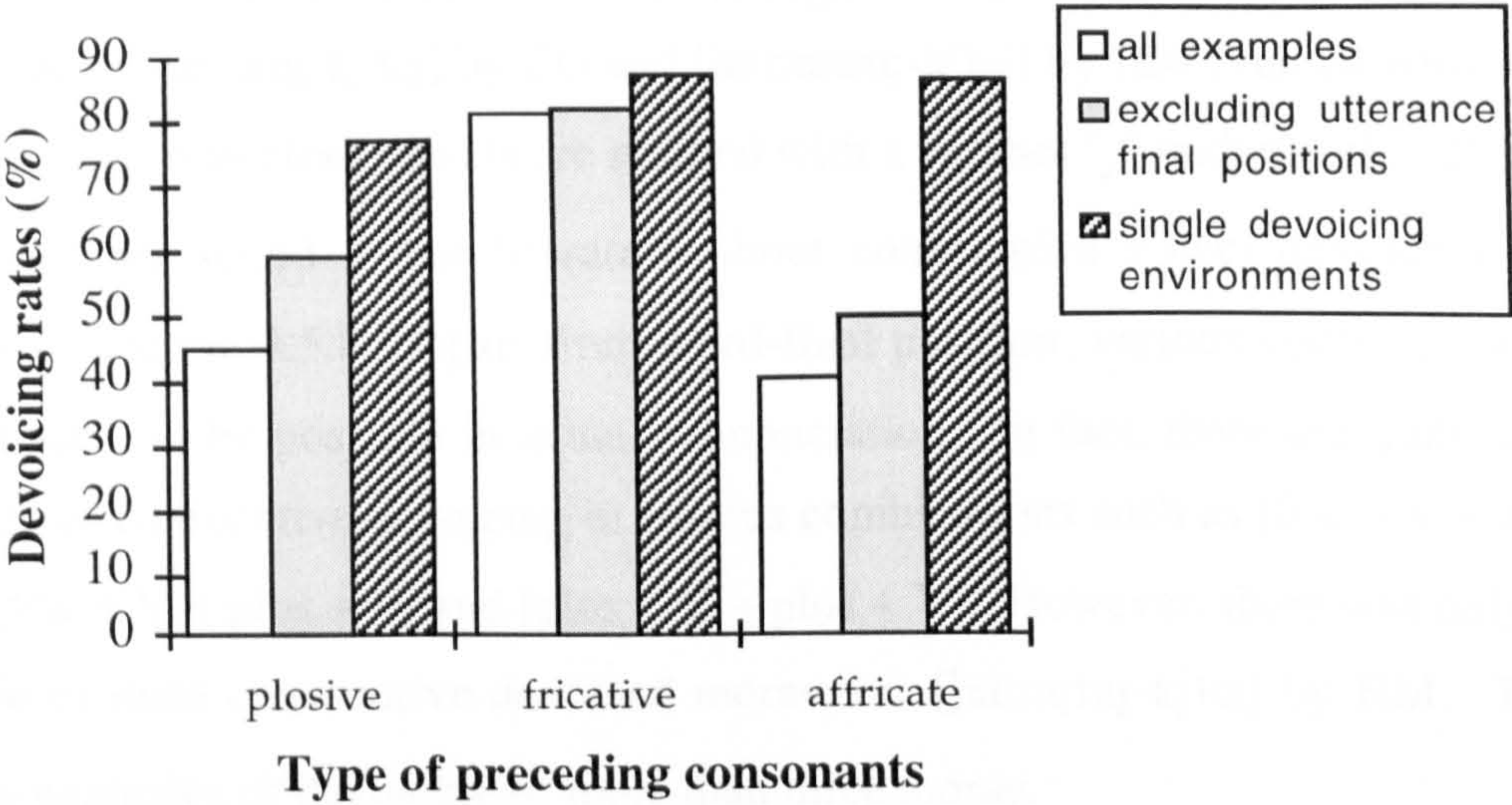
**Table 5.13      Average devoicing rates of examples with a single devoicing site by the type of preceding consonants**

	plosive	fricative	affricate	Average
devoicing rates	77.2 %	87.6 %	86.7 %	83.6 %
(No.)*	(125/162)	(155/177)	(91/105)	(371/444)

\*(the number of devoiced vowels/total number of samples)



**Figure 5.11 Comparison of average devoicing rates of all examples, examples excluding utterance final positions, and examples excluding utterance final positions and consecutive devoicing environments by the type of preceding consonants**



When examples in utterance final position and examples with consecutive devoicing environments were excluded, the devoicing rates were much higher than the rates in Tables 5.6 and 5.12. The devoicing rates of examples preceded by fricatives were higher than those of examples preceded by plosives and affricates in all positions and all conditions. A statistical analysis by a two-way ANOVA suggested that there was significant difference in the devoicing rates between examples with single devoicing environments and examples with consecutive devoicing environments [ $F(2,22) = 5.569, p < .025$ ], and that the difference between three types of preceding consonants was also significant [ $F(2,22) = 10.105, p < .001$ ]. However, the interaction between devoicing condition and type of preceding consonants was insignificant [ $F(4,22) = 1.057, n.s.$ ].



#### 5.4.7 The effect of the number of blocking factors on devoicing rates

373 examples with 881 devoicing sites were examined according to the presence of proposed blocking factors of vowel devoicing, and the results are listed in Tables 5.14a ~ 5.14c. First of all, most word-final high vowels remained voiced except for two examples [tɕo:φɯ-kʲɪ̥tɕi] by CO and [jʷɕʷɪtsʷ-kʲɪ̥tɕi] by HM (voiced vowels are underlined and voiceless vowels are marked with a symbol "̥" underneath). Despite what has been stated in the literature about consecutive vowel devoicing (see Chapter 4, Section 4.5.8.), apart from word-final position, various combinations of voicing seem to be possible in actual pronunciation. In fact, there are quite a few instances of consecutive devoicing in various combinations such as [fric + V̥ + affric + V̥], [fric + V̥ + plos + V̥] and [plos + V̥ + plos + V̥]. However, there was only one example of three consecutive devoiced morae, i.e. [jʷɕʷɪtʷ-kʲɪ̥tɕi] by HM. There were no examples of devoicing of more than three morae.

The experimental results suggested that devoicing rates were very high in the single devoicing environment, but not very high in the consecutive devoicing environment. This suggests that the blocking factors had little effect on devoicing rates in the single devoicing environment. However, in the consecutive devoicing environment, the devoicing rates were not necessarily high. Not all devoiceable vowels were devoiced.

Tables 5.14 (a), (b) and (c)

Patterns of vowel devoicing in the consecutive devoicing environment

- N.B. (i) The total number of samples for each test word is 18, except /desaki-kikaN/, /kiNki-tihoo/, /kiri huki-ki/ for which the total numbers are 17, and for /tukusi-tetudoo/ the number is 16.  
(ii) Devoiced vowels are indicated by the symbol of [◌◌] under the vowels.  
(iii) Voiced vowels are underlined.  
(iv) [ŋ] means a moraic unreleased nasal (/N/).  
(v) For gloss, refer to Tables 5.1a and 5.1b.

(a) Devoicing sites over two morae

test words	No.
[garasʊ-ki]	18

[ɕo:ɕi̥-tɕi]	11
[ɕo:ɕi̥-tɕi]	7

[ho:tɕi̥-ki]	15
[ho:tɕi̥-ki]	3

[do:kʊtʊ]	16
[do:kʊtʊ]	2

[ɸʊkʊ-toʔɕiŋ]	12
[ɸʊkʊ-toʔɕiŋ]	6

[aitɕi-tɕi̥ho:]	13
[aitɕi-tɕi̥ho:]	5

[kʊtsʊ-koʔdzo:]	11
[kʊtsʊ-koʔdzo:]	7

[kiŋki-tɕi̥ho:]	10
[kiŋki-tɕi̥ho:]	5
[kiŋki-tɕi̥ho:]	2

test words	No.
[rjokʊtɕi-koʔ:en]	10
[rjokʊtɕi-koʔ:en]	6
[rjokʊtɕi-koʔ:en]	1
[rjokʊtɕi-koʔ:en]	1

[ɕiɸʊ-ka]	12
[ɕiɸʊ-ka]	6

[kasʊpi̥-kai]	17
[kasʊpi̥-kai]	1

[giɸʊ-tɕi̥ho:]	11
[giɸʊ-tɕi̥ho:]	7

[to:hokʊ-tɕi̥ho:]	10
[to:hokʊ-tɕi̥ho:]	8

[desakj-ki̥kaŋ]	9
[desaki-ki̥kaŋ]	5
[desakj-ki̥kaŋ]	2
[desaki-ki̥kaŋ]	1

(Table 5.14 continues)

(Table 5.14 continued)

(b) Devoicing sites over three morae

test words	No.
[kok <u>u</u> h <u>u</u> k <u>u</u> ]	18

[k <u>u</u> t <u>s</u> u <u>ɕ</u> i <u>ta</u> -ko':dzo:]	18
--	----

[ts <u>u</u> k <u>u</u> ɕi-te'ts <u>u</u> do:]	13
[ts <u>u</u> k <u>u</u> ɕi-te'ts <u>u</u> do:]	3

[t <u>ɕ</u> o:h <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	7
[t <u>ɕ</u> o:h <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	6
[t <u>ɕ</u> o:h <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	3
[t <u>ɕ</u> o:h <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	1
[t <u>ɕ</u> o:h <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	1

test words	No.
[kiri <u>ɸ</u> u <u>k</u> i'-k <u>i</u> ]	17

[bok <u>u</u> t <u>ɕ</u> i <u>k</u> <u>u</u> ]	8
[bok <u>u</u> t <u>ɕ</u> i <u>k</u> <u>u</u> ]	6
[bok <u>u</u> t <u>ɕ</u> i <u>k</u> <u>u</u> ]	2
[bok <u>u</u> t <u>ɕ</u> i <u>k</u> <u>u</u> ]	2

(c) Devoicing sites over four morae

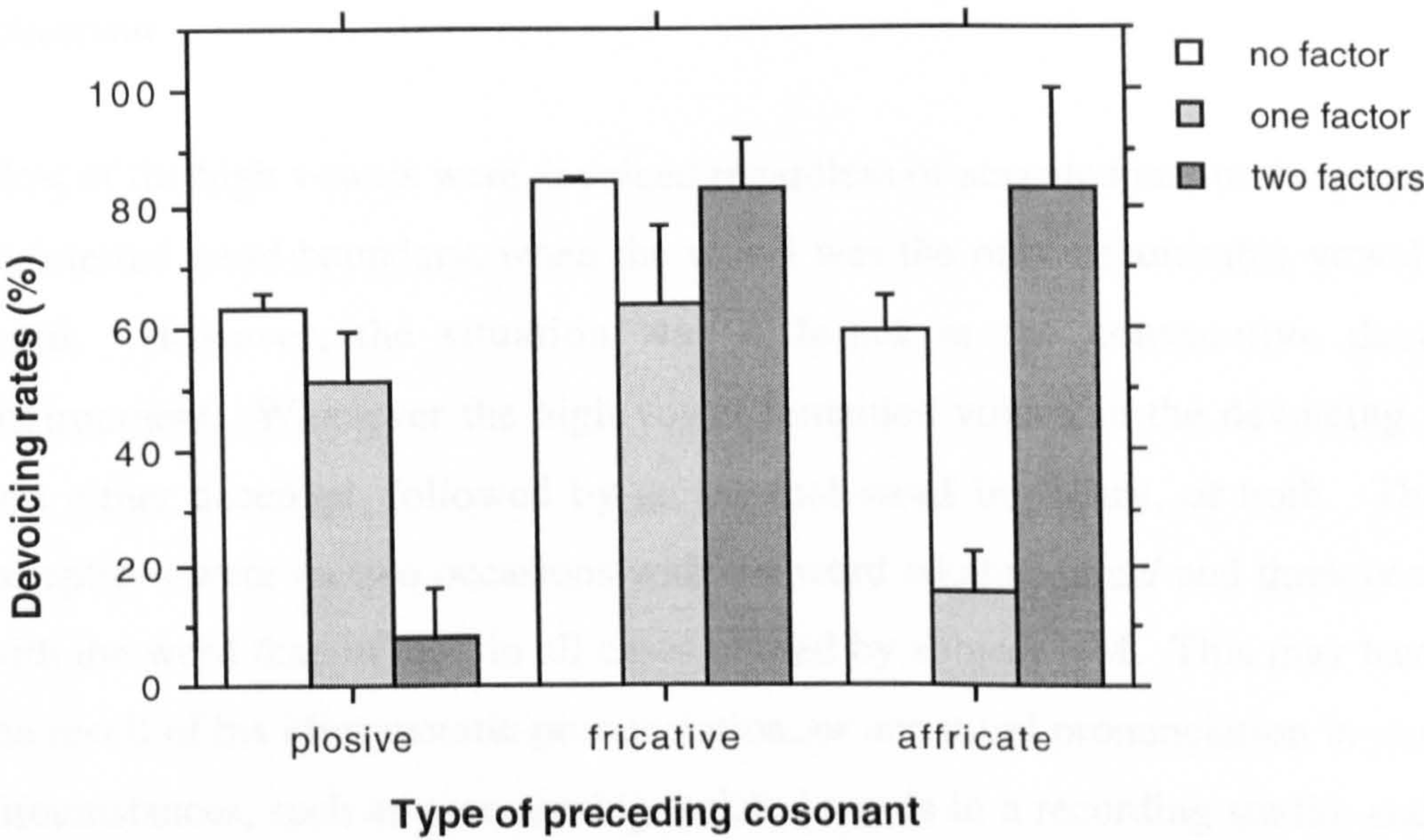
test words	No.
[j <u>u</u> ɕ <u>u</u> t <u>s</u> <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	8
[j <u>u</u> ɕ <u>u</u> t <u>s</u> <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	7
[j <u>u</u> ɕ <u>u</u> t <u>s</u> <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	1
[j <u>u</u> ɕ <u>u</u> t <u>s</u> <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	1
[j <u>u</u> ɕ <u>u</u> t <u>s</u> <u>u</u> -k <u>i</u> 't <u>ɕ</u> i]	1

To investigate whether blocking factors and the number of the blocking factors have any effect on the devoicing rates in the consecutive devoicing environment, the data were analysed by a two-way ANOVA taking the number of blocking factors and the type of preceding consonant as factors. The statistical results suggested that the effect of both the number of factors and the type of preceding consonants were significant: the result of the number of blocking factors was [F(2, 20), p<.025], and



the result of the type of preceding consonants was [F(2, 20),  $p < .0001$ ]. The interaction between these factors was also significant [F(4, 20),  $p < .0001$ ]. The results were shown in Figure 5.12.

**Figure 5.12**    **Average devoicing rates by the number of blocking factors and the type of preceding consonant in the consecutive devoicing environment**  
( | = Standard error of means)



Devoicing rates were generally higher for the preceding fricative. It was only when vowels were preceded by plosives that the highest devoicing rate occurred with no blocking factor and the lowest devoicing rate with two blocking factors. When the vowel was preceded by fricatives and affricates, the presence of two blocking factors did not necessarily block devoicing.

The results of devoicing rates as affected by the number of factors. Although the effect of the number of blocking factors was significant, a further analysis suggested that there was not a consistent reduction in the devoicing rate as the number of



blocking factors increased. In fact, for the vowels preceded by affricates, the devoicing rate was the highest when there were two blocking factors.

Also for the examples preceded by fricatives, the devoicing rate with two blocking factors was almost as high as those with no factor. The devoicing rate was highest when there was no blocking factor present. However, the devoicing rate was higher when there were two blocking factors rather than only one factor. This suggests that it was not the number of blocking factors which determined the devoicing rate, but rather the presence or absence of blocking factors and the type of preceding consonant.

Most of the high vowels were devoiced regardless of accentedness or the presence of an internal word boundary, when the vowel was the only voiceable vowel in the word. However, the situation was different in the consecutive devoicing environment. Whenever the high vowel remained voiced in the devoicing site, it was either accented, followed by an internal word boundary, or both. The only exceptions were on two occasions with the word /akatuti-jama/ and three occasions with the word /kaputi'ino/, in all cases uttered by subject NM. This may have been the result of his idiosyncratic pronunciation, or unnatural pronunciation in unnatural circumstances, such as pronouncing isolated words in a recording studio, and as far as the cases of /kaputi'ino/ are concerned, it may have been simply because he pronounced the word in an authentic Italian way since he is a fluent Italian speaker who goes to Italy very often. The overall devoicing rates of subject NM were lower than the average of all subjects (see Tables 5.4a ~ 5.4f and Figure 5.9.).

## 5.5 Conclusions

The results of this experiment show that when there is only one devoicing site in a word, the high vowels /i/ and /u/ are almost always devoiced regardless of their position in a word (except for utterance-final positions) and their accentual status. Acoustical analysis of the vowel in the typical devoicing environment found that

when the vowel was not devoiced, neither was it fully voiced throughout its duration. Voicing sometimes ceased before the end of the vowel, and a delay in voicing was observed in some samples.

There is also physiological evidence. As mentioned in Chapter 4, Sawashima (1971) found that there was a positive effort of widening of the glottis during the production of devoiced vowel segments. A sequence of voiceless consonants with an intervening high vowel generally shows only one long glottal opening, which implies that the devoicing is a phonological process to be described by a phonological rule rather than merely skipping voiced phonation between voiceless consonants.

However, in the case of words where consecutive devoicing is possible, the phonological constraints of accentuation and internal word boundary become significant as do the phonetic condition of the type of adjacent consonants. Preceding fricatives definitely triggered devoicing more than preceding plosives and affricates, but the number of phonological blocking factors was not significant. Rather it was the presence or absence of a blocking factor that was important.

There were no obvious relationships between the devoicing patterns in consecutive morae, and the type of the blocking factors and preceding consonants. However, the following constraints were recorded: (1) when there was no blocking factor in a word, voiced high vowels in consecutive devoicing sites were not favoured, but they were allowed if a blocking factor is present, (2) devoicing of more than three consecutive morae was not favoured. The results suggest that devoicing is phonetically a natural process when short vowels occur between voiceless consonants, but this phonetically natural process is blocked by the presence of phonological factors.

As illustrated in Figure 5.8, vowel devoicing in Japanese might be, in fact, a gradual change between fully voiced vowels and completely voiceless vowels. This prediction may explain the optionality of devoicing in the typical devoicing environment and partial devoicing of some devoiceable vowels. This prediction



implies that Japanese vowel devoicing is similar to the processes of vowel reduction found in other languages such as English, German, Greek, etc. discussed in Chapter 2. Whether Japanese vowel devoicing involves any nature of the vowel reduction process will be examined in Chapters 7 and 8.

Apart from the blocking factors used in this experiment, there are other factors which are considered as crucial, but not chosen for the experiment, such as word frequency, and psychological heaviness (see Chapter 2). The devoicing rates obtained in the experiment may have been influenced by these hidden factors.

The important factor for vowel devoicing was not the number of blocking factors, but actually the presence or absence of blocking factors. The problem remains. If the number of factors is not important for the realisation of devoicing, why are these blocking factors sometime effective and sometimes not? There may be other factors such as the syllable structure which control devoicing processes in Japanese. This point will also be discussed in Chapter 8.

## Chapter 6

# The effect of different speaking rates on vowel devoicing

### 6.1 Introduction

In Chapter 2, the phenomenon called vowel devoicing, and similar phenomena such as vowel reduction and deletion in various languages were discussed. In most languages, the phonetically weak vowel [ə], or intrinsically short vowels such as high vowels and [ə], are often devoiced in an unstressed position, and vowel devoicing is usually treated as an optional process in fast or casual speech. In Standard Japanese, however, high vowel devoicing is treated as a compulsory phonological rule, because when the high vowels are voiced in the devoicing environment, it does not sound natural: it is a sign of foreign accent or non-standard dialect (Hirayama, 1985: 44, Maekawa, 1987: 143, etc., see Chapter 4). Chapter 5 demonstrated that high vowel devoicing occurred very regularly and confirmed that Japanese vowel devoicing was almost a compulsory process, but this process was limited to the environment where there was no devoiceable vowel in a neighbouring mora (the single devoicing environment). The claimed blocking factors such as accentedness, presence of word boundary and type of preceding consonant were not effective in the single devoicing environment. However, not all devoiceable vowels were devoiced in the environment where consecutive devoicing was possible. In the consecutive devoicing environment, these blocking factors became effective and blocked some of the devoiceable vowels from being devoiced.

Similar to other languages, vowel devoicing in Standard Japanese involves the high vowels /i, u/, which are intrinsically shorter in duration compared with other vowels, and the devoicing occurs between voiceless consonants or between a voiceless consonant and a pause, for both of which the glottis is wide open. In many languages, the process occurs particularly in fast or casual speech, in which the process initially involves vowel centralisation and durational reduction. However, Japanese vowel devoicing operates in different conditions from other languages. It is because (1) vowel devoicing in Japanese does not involve obvious vowel centralisation or durational reduction depending on stress or accentual status, as in English, German, Greek, Hebrew (see Chapter 2), (2) vowels are not normally neutralised in an unaccented position or in fast speech, and (3) devoicing occurs regularly not only in fast speech but also at a normal speaking rate and even in citation form (see Chapter 4). Physiological studies also reported that there was a positive movement of opening the glottis during the production of voiceless vowels (Sawashima, 1971 and Waitzman et al., 1976), unlike voicing of an intervocalic /h/, which showed clear signs of glottal assimilation with neighbouring (voiced) vowels and did not seem to involve any positive adductive movements of laryngeal muscles (see Chapter 4, Section 4.3). Both studies concluded that voicing of /h/ seemed to be a result of inevitable glottal assimilation, but vowel devoicing in Japanese was not just a mere skipping of phonatory activity but *was* a more positively controlled process by the speaker's intention.

Hasegawa (1979) made a crucial distinction between "casual speech" and "fast speech" based on whether their processes were controlled only by speech rate. She defined (1979: 126) "fast speech" as having processes which were conditioned purely by phonetic factors, remained neutral to sociological factors and were not lexically governed, and defined "casual speech" as having processes which were more or less sensitive to sociological notions and lexical information. Casual speech involves not only the characteristics of the phonological process but also those of morphological and syntactic processes. Hasegawa claimed that high vowel devoicing was a process in fast speech, because it applied across a word boundary or syntactic configurations but did not necessarily alter them.



Hasegawa added that in slow or careful speech, however, the high vowel devoicing process involved various factors such as accent, word boundary and syntactic configuration. Hasegawa's statement on this particular point is not entirely correct since she mixed slow speech and careful speech despite the fact that she was trying to differentiate two different notions of speech: slow speech is a matter of speech rate but careful speech is a matter of speech style. In addition, high vowel devoicing frequently occurs at a normal speaking rate (see Chapter 4, Section 4.4), and does not seem to be a particularly distinctive process of fast speech. Moreover, devoicing can eventually cause syntactic or morphological changes. For instance, female names /ki`kuko/ and /a`kiko/ are normally pronounced with voiceless [ɸ] and [j] as [ki`kɸko] and [a`kɸko] respectively in normal speech. In casual speech, however, these names are often pronounced as [ki`k:o] and [a`k:o] respectively with geminate consonants /kk/ as a result of high vowel deletion following devoicing. Despite some incorrectness, her claim was worth taking into account because she was right (1) to differentiate two different processes, (2) to mention that the vowel devoicing process was sensitive rather to speaking rate than to speech style, as it normally occurs in formal as well as casual speech.

Vowel devoicing in many languages may be explained as a case of phonetic undershoot or glottal assimilation. In speech, coarticulation or assimilation of neighbouring sounds is inevitable. The most common environment for vowel devoicing or reduction in many languages is in unstressed syllables. In many languages, everything being equal, the duration of unstressed vowels tends to be shorter than stressed vowels, and it is a language universal that voiceless consonants are generally longer than their voiced counterparts. When a short vowel occurs adjacent to voiceless consonants, the vowel duration may be too short for the vocal folds to start vibrating after a voiceless consonant, and then stop vibrating again for the following voiceless consonant. This is a physiologically oriented inevitable process rather than a phonologically controlled rule. Therefore, this process occurs more frequently as speaking rate increases.

Lindblom (1983) explained the effect of speaking rates on the movement of the articulatory organs in Swedish, using data on the movement of the tongue and jaw when the speaking rate was increased. Swedish vowel reduction in CVC syllables showed that formant frequencies tended to vary systematically as a function of vowel duration and consonant context, and explained the vowel reduction with relation to the formant 'undershoot' theory of speech production (Stevens and House, 1963). Creating a model of speech production system made up of mass, damping and spring components, Lindblom suggested that if the mass, spring and damping were constant, then the response to the forcing function, and the extent of the undershoot depended exclusively on the duration of the event.

The undershoot theory suggests that as the duration of the vowel decreases, there will be less and less time to complete the approach to the vowel target, and the vowel gesture will be more and more reduced. In the formant domain the value observed for the vowel will show the influence of the adjacent consonants more and more strongly (Lindblom, 1983: 228).

However, a series of studies by Gay et al. (1974) and Gay (1978) found that undershoot was not an inevitable consequence of short duration. Gay et al. (1974) studied the effect of speaking rate (slow and fast) on the articulation of the consonants /p, w/ in combination with the vowels /i, a, u/, using electromyography and high speed full-face motion pictures for lip movements, and high speed lateral view X-ray films for tongue and jaw movements. The results showed different effects of speaking rate on labial consonants and vowels. For the production of labial consonants, the major effects of an increased speaking rate were an increase in the activity level of the muscles and an increase in the rate of movement of the lips, as apparent consequences of an increase in articulatory effort. However, the data for the tongue and jaw movements showed a tendency to undershoot. The results suggested an increase in speaking rate was accompanied by a decrease in displacement and a decrease in the activity level of the muscle. The undershoot was not, however, as suggested by Lindblom (1983), but because of a consequence of an overlap in the timing of commands of the muscle.

The electromyographic data for the tongue suggested that vowels seemed to have different targets for slow and fast speech. “For slow speech, and perhaps for stressed vowels, one given set of coordinates is aimed for, while for fast speech, where articulatory expediency or the constraints of decreased jaw displacement place additional demands on the mechanism, a different set of coordinates is aimed for” (Gay et al., 1974: 62).

Further spectrographic study by Gay (1978) on the relationship between speaking rate and vowel formant movements using four American English speakers as subjects found that the midpoint formant frequencies of the different vowels did not vary depending on speaking rate, but for fast speech the onset frequencies of second formant transitions were closer to their target frequencies while CV transition rates remained essentially unchanged. In other words, movement toward the vowel simply began earlier for fast speech. However, the data combining speaking rate and lexical stress suggested different control mechanisms for speaking rate and lexical stress. For stressed vowels in fast speech, only duration was reduced to a certain degree, but unstressed vowels, even if they were of the same duration as quickly produced stressed vowels, were reduced in overall amplitude, fundamental frequency and to some extent vowel quality. Vowel neutralisation toward [ə] in fast speech occurred only in unstressed vowels.

Kuehn and Moll (1976) studied the relationship between articulatory velocity and displacement in different speakers and different phonetic contexts. Their results found that articulatory velocity was influenced by phonetic context and that different speakers used different physiological methods of changing speaking rates. As for phonetic context, everything being equal, velocity movements both towards and from fricatives were produced with a slower movement than stops because the fricatives are characterised by vocal tract constrictions of carefully controlled and relatively constant size. With relation to speaking rate, when speaking rate was increased, some speakers did not change velocity or actually reduced velocity, and instead decreased displacement, whereas other speakers increased velocity so that the articulatory position was achieved for the phone at a slow rate. Kuehn and Moll concluded that “in



order to accomplish various magnitudes of displacement within relatively fixed periods of time, it would appear that the fixed time period and the eventual position (target) that the articulator should reach must be known to the production system and that some mechanism is required which computes necessary velocity and controls movement accordingly" (1976: 318).

Lindblom (1983) concluded that the target achievement of articulatory organs was dependent on two factors, i.e. segment duration and vocal effort. The speech production system is capable of raising the level of its performance, but it prefers not to. "Speech production appears to operate as if physiological processes were governed by a power constraint limiting energy expenditure per unit time" (Lindblom, 1983: 231).

There has been little research on the effect of speaking rate on vowel devoicing. Maekawa (1990) carried out two experiments on the effect of speech rate on the following three: (a) devoicing of the high vowel /u/ in the CVC environment (the following consonant is voiced), (b) devoicing of the non-high vowel /a/ in the CVC environment, and (c) voicing of intervocalic /h/.

In his first experiment, Maekawa used nonsense words controlling various combination of consonants adjacent to a high vowel /u/ at two speaking rates, normal and fast. Maekawa categorised vowels into 3 voicing states: (i) showing clear vowel formants for a certain duration on the spectrogram, (ii) no clear formants observed on the spectrogram, but there were very short waves (usually only one cycle) observed on the waveform, and (iii) either clearly voiceless or only aperiodic waves observed on the waveform.

Maekawa's results showed that when only the voicing state of vowels (i) above was considered, the /u/-devoicing<sup>1</sup> was very rare at the normal rate, and varied depending

---

<sup>1</sup> In his paper, Maekawa used the term '/u/-weakening' meaning /u/-devoicing in the CVC environment, in order to distinguish the phenomenon from normal /u/-devoicing in the CVC environment. However, I use the term 'devoicing' here in order to differentiate vowel weakening phenomenon which will be discussed in Chapter 7.

on the type of preceding consonants at the fast rate: almost none when preceded by voiceless plosives but about 25 ~ 35% when preceded by fricatives or affricates. However, when examples of voicing state (ii) above were included, the /u/-devoicing occurred about 30 % at the normal rate when preceded by fricatives or affricates, and the /u/-devoicing rates were higher (45 ~ 65%) for all types of preceding consonants at the fast rate.

In terms of non-high vowel /a/-devoicing, devoicing rates were much higher at the fast rate than the normal rate, but the results varied depending on preceding consonants and positions in a word. When the vowel /a/ was preceded by /h/, there was a difference in its devoicing rates depending on its position in a word at the normal rate: about 40% in the word initial position but almost none in the word internal position. On the other hand, at the fast rate, very little difference was observed between word initial and word internal positions (about 80% and 70% respectively). When the vowel /a/ was preceded by /s/, devoicing rates were lower than when preceded by /h/ in both word initial and internal positions at the normal rate. Devoicing rates at the normal rate were 0% in a word internal position and 30% in word initial position, but there was a big difference in the /a/-devoicing rates preceded by /s/ between the two positions at the fast rate; the devoicing rate was about 30% in word internal positions, but it was about 90% word initially.

In Maekawa's second experiment, he examined the same three voicing/devoicing phenomena at 7 different rates ranging from 'very slow' to 'extremely fast' with relation to the position in a word and the type of a vowel in the following syllable. In this experiment, Maekawa defined 'devoiced /u/' as the category (iii) above only, i.e. vowels which showed even a single pulse on their waveforms were not counted as devoiced. The results showed that /u/-devoicing occurred only at the two fastest rates. Devoicing of /u/ was also more frequent for both positions when the vowel in the following syllable was the non-high vowel /a/. Devoicing of /a/ and voicing of /h/ were dependent on their positions in a word. Voicing of /h/ in word initial position was more common even at fairly slow rates, but in word internal position it occurred

only at the fastest rate. As for the devoicing of /a/, devoicing rates were very high in word initial position, but less than 50 % in word internal position.

As for /u/-devoicing, even at normal tempo /u/ followed by a voiced consonant was reduced to a very brief, but periodic sound on a spectrogram in about 30% of all samples, but for /a/ to be devoiced in the above environment (b) a much faster tempo was required. Maekawa concluded that when a high vowel was preceded by a voiceless consonant, it was not necessary for the high vowel to be followed by a voiceless consonant in order to be reduced or devoiced. As far as the type of preceding consonants was concerned, both types of devoicing were more likely to occur when the vowel was preceded by a fricative [s] or an affricate [ts], and the reduction and devoicing rates were much lower when preceded by a plosive [k] at both normal and fast speaking rates. The devoicing of /a/ started to occur at rates slightly faster than normal. The devoicing rates of /a/ in the first syllable of a word were 100% at speaking rates which were faster than normal, but in word medial position it was less than 50% even at the fastest tempo.

Maekawa's results suggested that in Japanese high vowels following voiceless consonants were likely to be reduced even when the vowel was followed by voiced consonants and even at a normal speaking rate, but more commonly at a faster rate, but for the same vowels to be devoiced in non-devoicing environments, much faster tempo was required. Non-high vowel devoicing and /h/-voicing were strongly influenced not only by speaking rate but also by various factors such as position in a word and preceding consonants.

Maekawa's study implied that there were two devoicing processes in Japanese: one was influenced by speaking tempo and the other was not affected by speaking tempo. An interesting observation was that when the tempo increased, the vowel devoicing rate did not necessarily increase but vowel reduction rate increased. That probably implied that vowel devoicing is part of an overall process of vowel reduction, similar to the process in other languages. Both high vowel devoicing in non-devoicing environments and non-high vowel devoicing were sensitive to speaking rate and



occurred more often at faster rates. As they do not have ideal environments for vowel devoicing, a faster tempo was required in order to skip phonatory adjustment activities.

Vowel devoicing in other languages seems to be more prominent when speaking rate increases. When speaking rate increases, the articulation of peripheral vowels becomes centralised and approaches [ə], and their durations also become shorter, and eventually they become voiceless or in an extreme case are deleted. Vowel devoicing in Japanese seems to operate differently. Japanese does not have obvious vowel centralisation or durational reduction, but if Japanese vowel devoicing is a result of economy of speech gesture skipping vowel phonation between voiceless consonants, it should occur more at faster speech rates and less at slower rates.

In this chapter, I will examine whether Japanese vowel devoicing is in principle phonetically controlled; in other words, whether devoicing is directly affected by an increase in speaking rate. If devoicing is phonetically motivated, it should occur more at faster tempi than slower tempi. The process will be examined in terms of another phonetic factor, namely the type of preceding consonant in relation to a phonological factor, namely single and consecutive devoicing conditions.

## 6.2 Experimental hypotheses

Chapter 5 showed that Japanese vowel devoicing was almost a compulsory rule. Although the previous experiment in Chapter 5 and Maekawa (1990) demonstrated that non-high vowels could be devoiced between voiceless consonants and high vowels could also be devoiced when they were adjacent to a voiced consonant, it was predominantly high vowels that were devoiced between voiceless consonants. This is a phonetically ideal environment for devoicing: a short vowel surrounded by voiceless sounds for which the glottis is wide open. Vowel devoicing involving non-high vowels or adjacent to voiced consonants is rare in ordinary speech (Takeda & Kuwabara, 1987, Sakurai, 1985, Chapters 4 & 5). If vowel devoicing in Japanese is

a physiologically oriented process, similar to processes found in other languages, when the speech rate increases the vowel devoicing rates will also increase and voicing of vowels will be more often retained at a slow speaking rate. On the other hand, if devoicing is more of a phonologically controlled process, the effect of speaking rates will be minimal and devoicing will also occur at a slow speaking rate.

There are some linguistic factors which are sensitive to speaking rate. For example, under normal circumstances in speech, a speaker does not pause at an internal-word boundary, for example in a compound word, at a normal speaking rate, but sometimes inserts a pause at a slower rate. A speaker tends to slow down speech in utterance-final position (Sagisaka, 1984, Takeda et al., 1989, Kaiki et al., 1990). Earlier results presented in Chapter 5 showed that devoicing is rare in utterance-final position. If a speaker inserts a pause at an internal word boundary, the mora before the pause becomes utterance-final. Then if the same retardation process occurs, a devoiceable vowel followed by an internal word boundary is likely to remain voiced. Therefore, word boundary effects, such as insertion of a pause, might be more strongly influenced by speaking rate than other linguistic factors. On the other hand, if there are some blocking factors which are less sensitive to speaking rate, they may still be effective in fast speech and block vowels which undergo a devoicing process.

The hypotheses relevant to this present experiment are:

- (1) Japanese vowel devoicing is primarily a phonetic process caused by glottal assimilation with adjacent voiceless consonants, and it occurs more at a faster tempo than a slower tempo.
- (2) Different processes operate in vowel devoicing depending on its environment, i.e. single or consecutive, since single and consecutive devoicing environments show different devoicing rates.
- (3) Some blocking factors of vowel devoicing are sensitive to speaking rate, whereas others are not.

## 6.3 Experimental methods

### 6.3.1 Subjects and procedure

The same six people who took part in the previous experiment in Chapter 5 were the subjects for the present study. The recordings were done at the same session as for the previous experiment, but the subjects were asked to read the reading texts before they read the list of test words for the experiment in Chapter 5, and were given time to familiarise themselves with the texts. They were asked to read the texts at three different tempi: 'fast', 'normal' and 'slow', three times for each tempo. Each speaker read the text A; firstly three times at a 'normal' tempo, secondly three times at a 'fast' tempo, and thirdly three times at a 'slow' tempo. If they made any mistakes in reading, they were allowed to stop, correct and repeat the part as many times as they wished.

### 6.3.2 Materials

The test words used were the same as for the experiment described in Chapter 5: a total of 41 test words, 20 words with a single devoicing site and 21 words with consecutive devoicing sites. Devoiceable vowels were preceded by all types of possible voiceless consonants in order to see the effect of preceding consonant type, and they were always followed by a voiceless plosive, which helped to define the segmental boundary more clearly when the vowel was devoiced and made segmentation easier. Two reading texts were made up containing all the test words; (1) text A contains 20 test words and (2) text B contains the remaining 21 test words. Text A is a short paragraph explaining a route on a map. Text B is a short paragraph of a series of sentences which are grammatically correct but have no logical meaning (see Appendix IIa and IIb)

### 6.3.3 Speech rate setting

An experiment was carried out to examine vowel devoicing rates in a reading style at three different speaking rates. In order to examine the effect of speech rate on vowel



devoicing, three speaking tempi were chosen for the experiment: namely (a) ‘fast’, ‘normal’ and ‘slow’. The subjects were instructed to target the ‘normal’ tempo as a comfortable speaking rate, the ‘fast’ tempo as considerably faster than a comfortable rate but not as fast as to stutter or make many mistakes, and the ‘slow’ tempo as considerably slower than a comfortable rate but tokens spoken at this slow rate still sounded natural. They were asked to control the speaking tempi by themselves because comfortable speaking rates vary from speaker to speaker.

#### **6.3.4 Data analysis**

Recorded data were analysed using the ESPS Waves+ package at the sampling rate of 16 KHz. For each test word, a speech pressure waveform and a spectrogram were created to examine the presence or absence of voicing. The voicing of vowels was decided by the same criteria as in the experiment in Chapter 5: primarily from the speech pressure waveform, and secondarily from the spectrogram. Vowels were categorised into two groups, namely ‘voiced’ and ‘voiceless’ vowels. Partially voiced vowels (for their definition, see Chapter 5) were included in ‘voiced’ vowels, but ambiguous vowels were included in ‘voiceless’ vowels.

### **6.4 Experimental results and discussion**

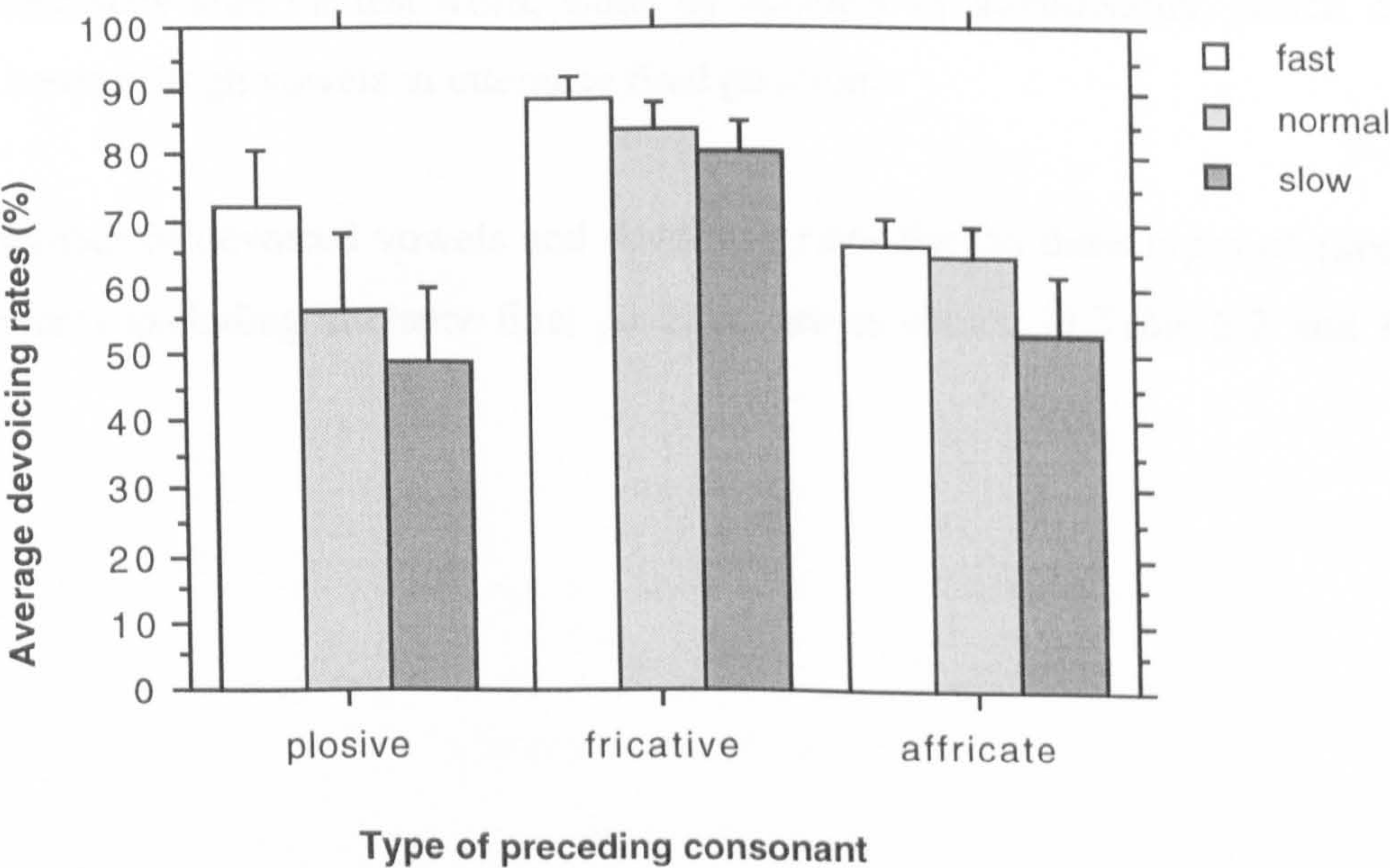
#### **6.4.1 Effect of the type of preceding consonants**

Firstly devoicing rates at the three tempi were calculated by the type of preceding consonants, and are shown in Table 6.1 and Figure 6.1. The results showed that devoicing rates were higher at the faster tempi and lower at the slower tempi for all preceding consonant types as expected.

**Table 6.1** The total number of devoiced vowels of 6 subjects determined by the type of preceding consonant at three tempi  
(Number of devoiced vowels/total number of examples)

Type of consonant \ Tempo	Fast	Normal	Slow
Plosive	340/465 examples (73.12%)	284/465 examples (61.8%)	262/466 examples (56.22%)
Fricative	329/377 examples (87.27%)	313/377 examples (83.02%)	289/376 examples (76.86%)
Affricate	212/324 examples (65.43%)	206/324 examples (63.58%)	166/324 examples (51.23%)

**Figure 6.1** Devoicing rates determined by the type of preceding consonant at three tempi (l = Standard error of means)





The results of a two-way analysis of variance (ANOVA) showed that the effect of tempo was not significant [ $F(2,24)=2.315$ , n.s.]. However, the effect of consonant types was significant [ $F(2,24)=11.284$ ,  $p < .001$ ]. Devoicing rates were higher when the vowels were preceded by fricatives than by affricates and plosives. No significant interaction between speaking tempo and consonant types was observed [ $F(4,24)=.363$ , n.s.]. The results were similar to the results of the previous experiment in Chapter 5: devoicing is most common when the high vowels are preceded by voiceless fricatives and least common when preceded by voiceless affricates.

#### 6.4.2 The effect of position in an utterance

The previous experiment in Chapter 5 suggested that vowel devoicing was less common in utterance final position. The texts for this experiment were prose, and no test words happened to appear in the sentence or phrase final position, except for one word *kirihukiki* 'sprayer' which is in the position typically followed by a pause because the word was followed by a boundary between two juxtaposed phrases (see Appendix IIb). There were quite a few cases when a subject paused in the middle of, or immediately after the test word, either by accident or intentionally, which created some cases of high vowels in utterance final position.

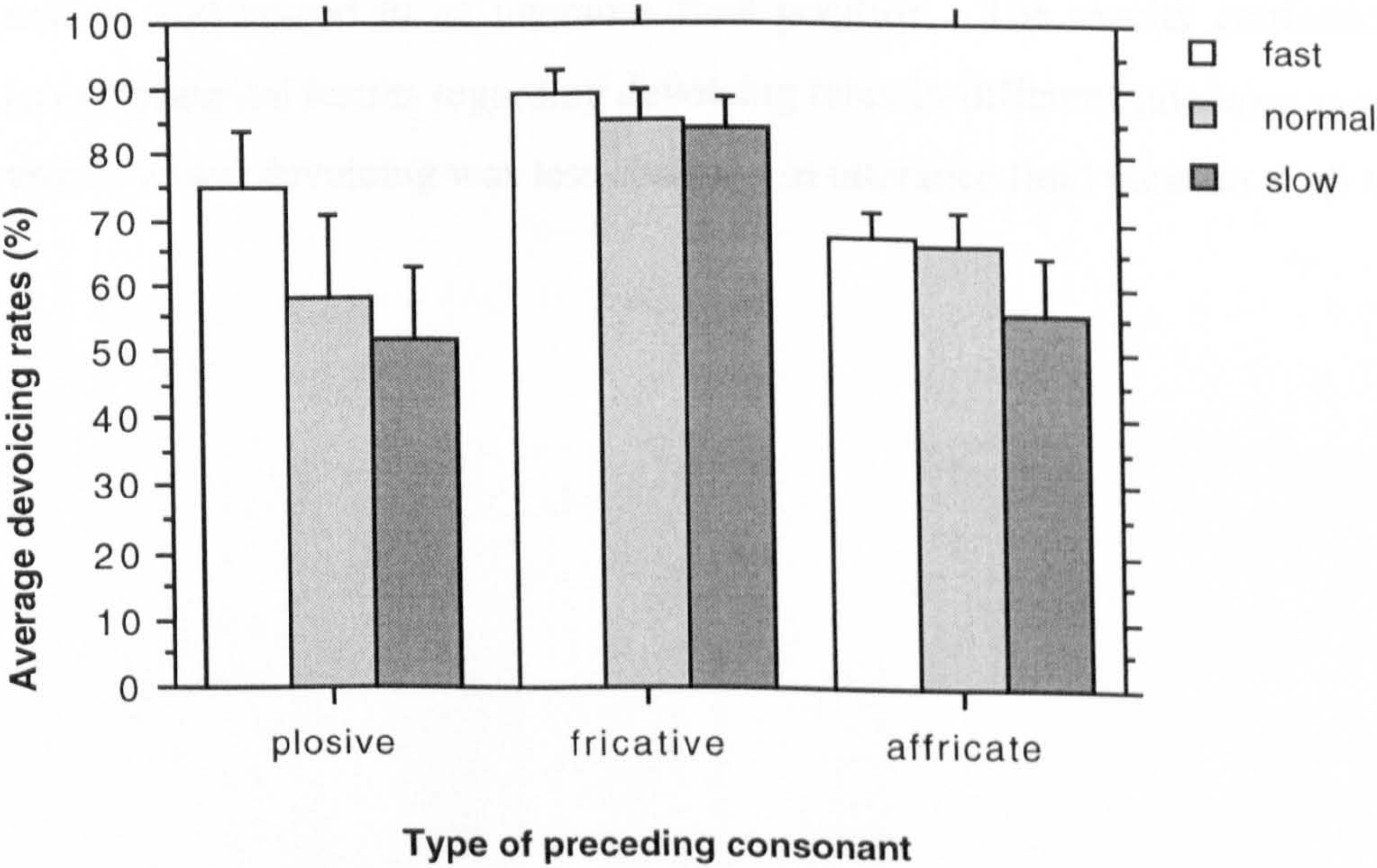
The number of devoiced vowels and devoicing rates for the three types of preceding consonants excluding utterance final positions are as shown in Table 6.2 and Figure 6.2.



**Table 6.2    The total number of devoiced vowels of 6 subjects determined by the type of preceding consonant excluding utterance final positions at three tempi (Number of devoiced vowels/total number of examples)**

<div>Type of consonant \ Tempo</div>	Fast	Normal	Slow
Plosive	339/446 examples (76.01%)	284/445 examples (63.82%)	262/441 examples (59.41%)
Fricative	326/370 examples (88.11%)	313/371 examples (84.36%)	288/350 examples (82.29%)
Affricate	212/320 examples (66.25%)	206/318 examples (64.79%)	166/309 examples (53.72%)

**Figure 6.2    Devoicing rates determined by the type of preceding consonant excluding utterance-final positions, at three tempi (I = Standard error of means)**





The devoicing rates for all positions at all tempi were slightly higher compared with the devoicing rates including utterance-final positions in Figure 6.1, as shown in Figures 6.3a, 6.3b and 6.3c. In this condition, too, the mean devoicing rates were larger at the faster rates than the slower rates for all types of preceding consonant as when the data included utterance-final positions.

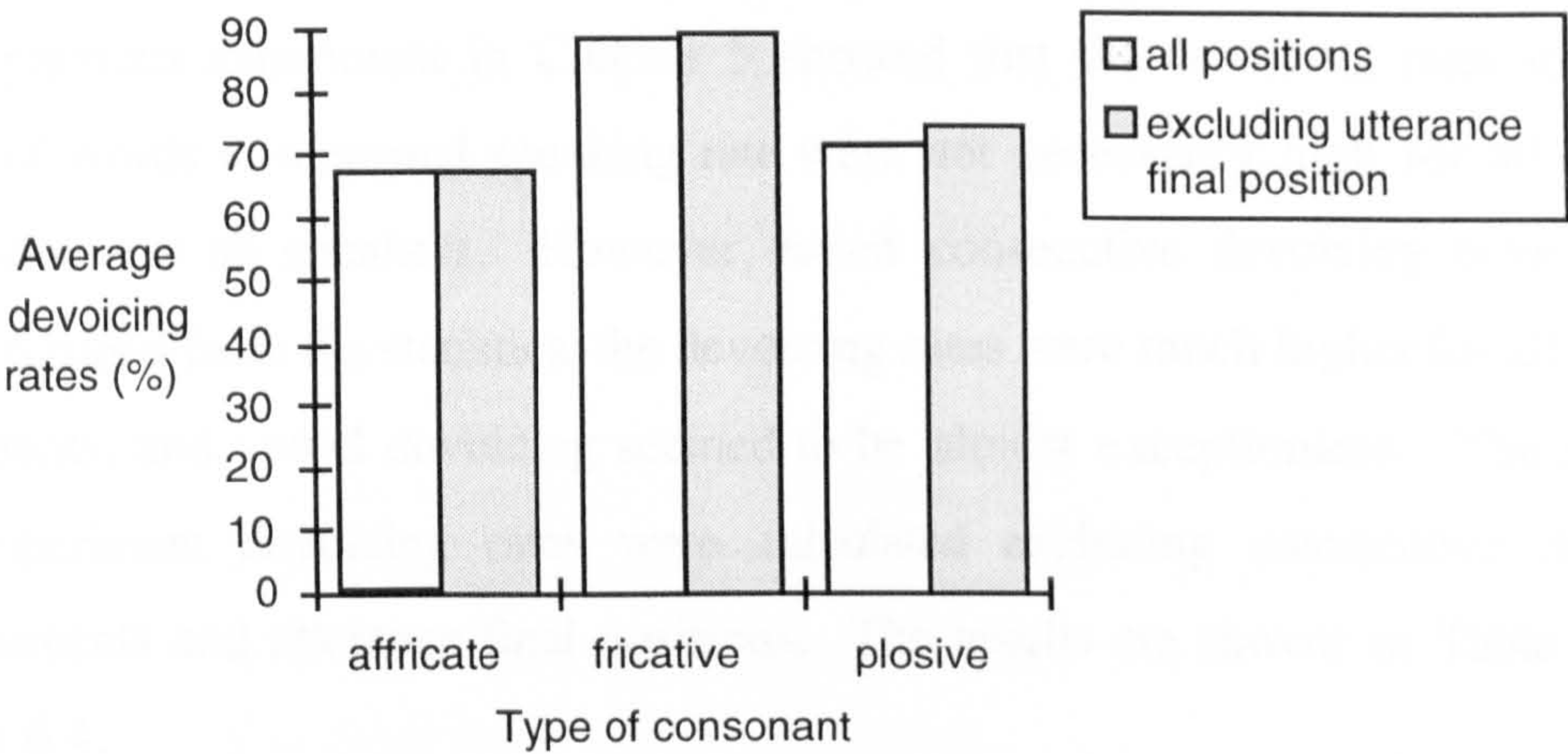
The results of a two-way ANOVA showed that there was no effect of tempo on the devoicing rates [ $F(2,24)=1.780$ , n.s.] for this devoicing condition either, but the effect of consonant types was again significant [ $F(2,24)=10.258$ ,  $p < .001$ ]: preceding fricatives trigger significantly more devoicing than preceding plosives and affricates. The interaction between speaking tempo and consonant types was not significant [ $F(4,24)=.458$ , n.s.].

There were quite a few examples where the subjects inserted a pause either between a test word and the following word, or at an internal word boundary. This tendency was more common at the slower tempi than the faster tempi. When the subjects did not insert a pause in one utterance, but did insert a pause in another utterance of the same test word, the devoiceable vowels in question were usually voiceless in a non-final position and voiced in an utterance final position. The results confirmed the previous experimental results regarding devoicing rates in different utterance positions in Chapter 5: vowel devoicing was less common in utterance final position at all tempi.

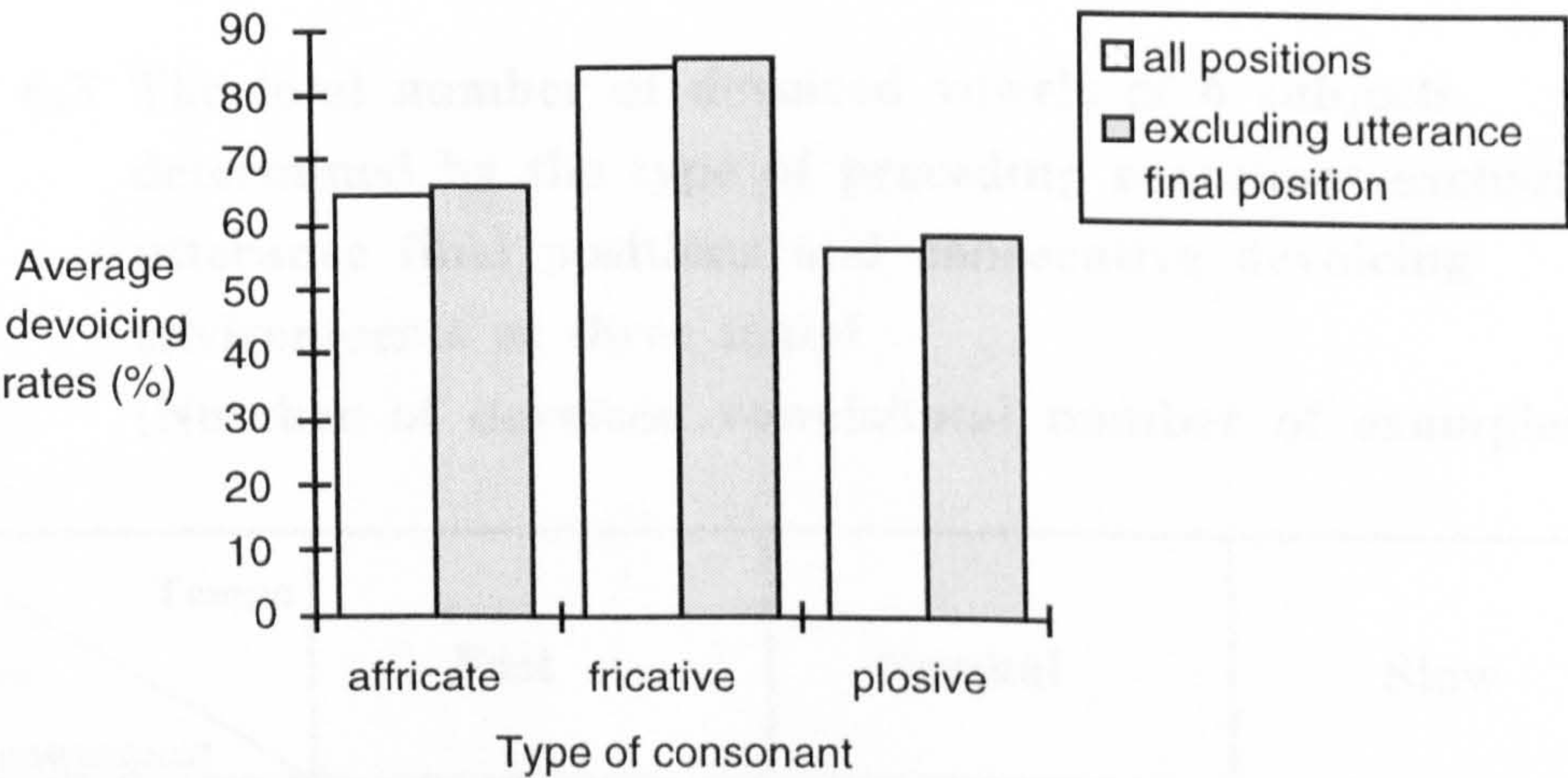
Figures 6.3a, 6.3b and 6.3c

Comparison of the devoicing rates at three tempi by the types of preceding consonants in two devoicing conditions

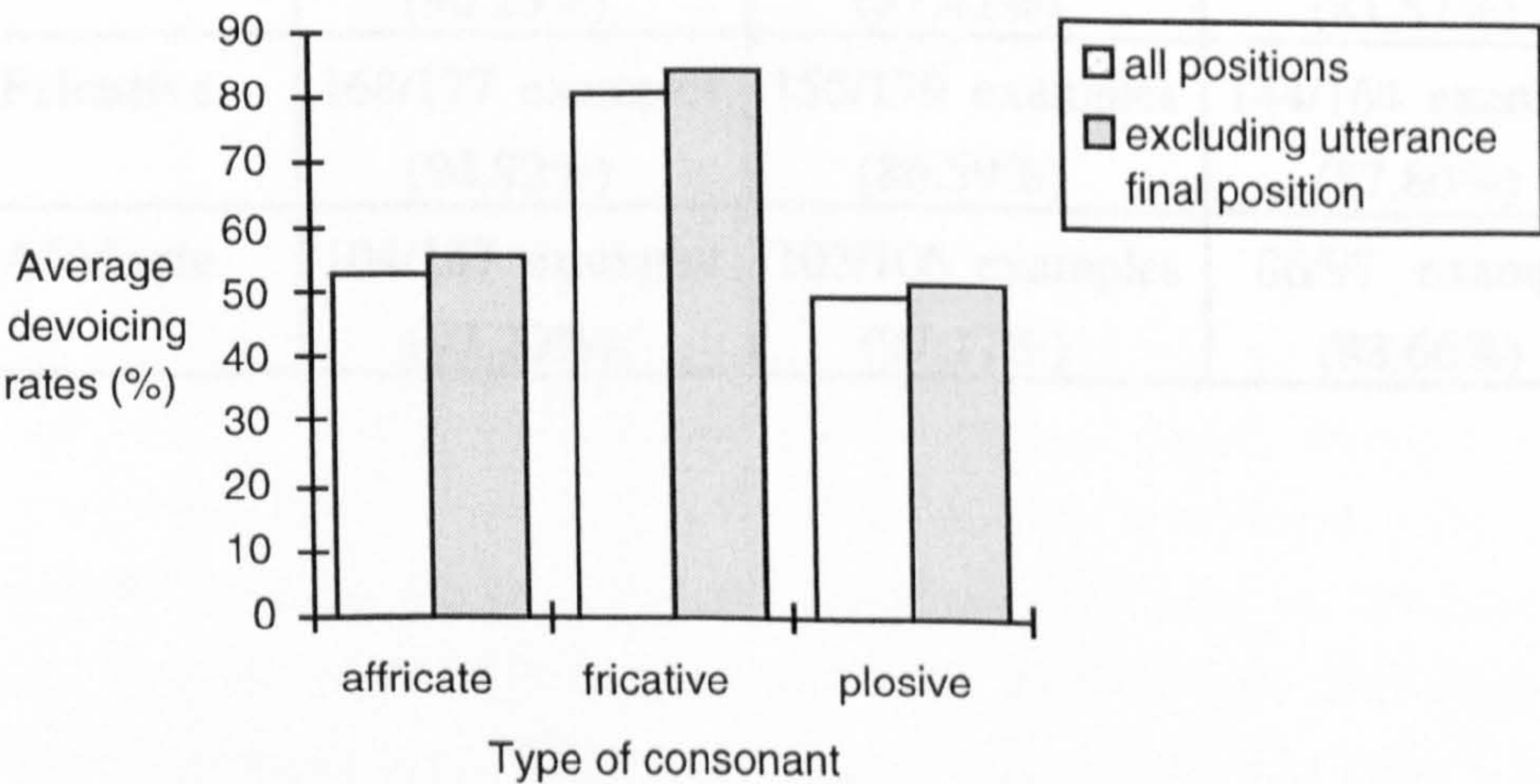
(a) Fast tempo



(b) Normal tempo



(c) Slow tempo





6.4.3 Effect of devoicing conditions

Another important factor for vowel devoicing rates in the previous experiment was devoicing conditions, i.e. single or consecutive devoicing environments. The results of the previous experiment in Chapter 5 showed that the devoicing rates in citation forms of words at a normal speaking rate were not necessarily high for all types of consonants for all speakers. However, when consecutive devoicing environments were excluded from the statistics, the devoicing rates were much higher for all types of consonants, and vowel devoicing seemed to be almost exceptionless. Therefore, in this experiment devoicing rates were calculated excluding consecutive devoicing environments and utterance final positions. The results are shown in Table 6.3 and Figure 6.4.

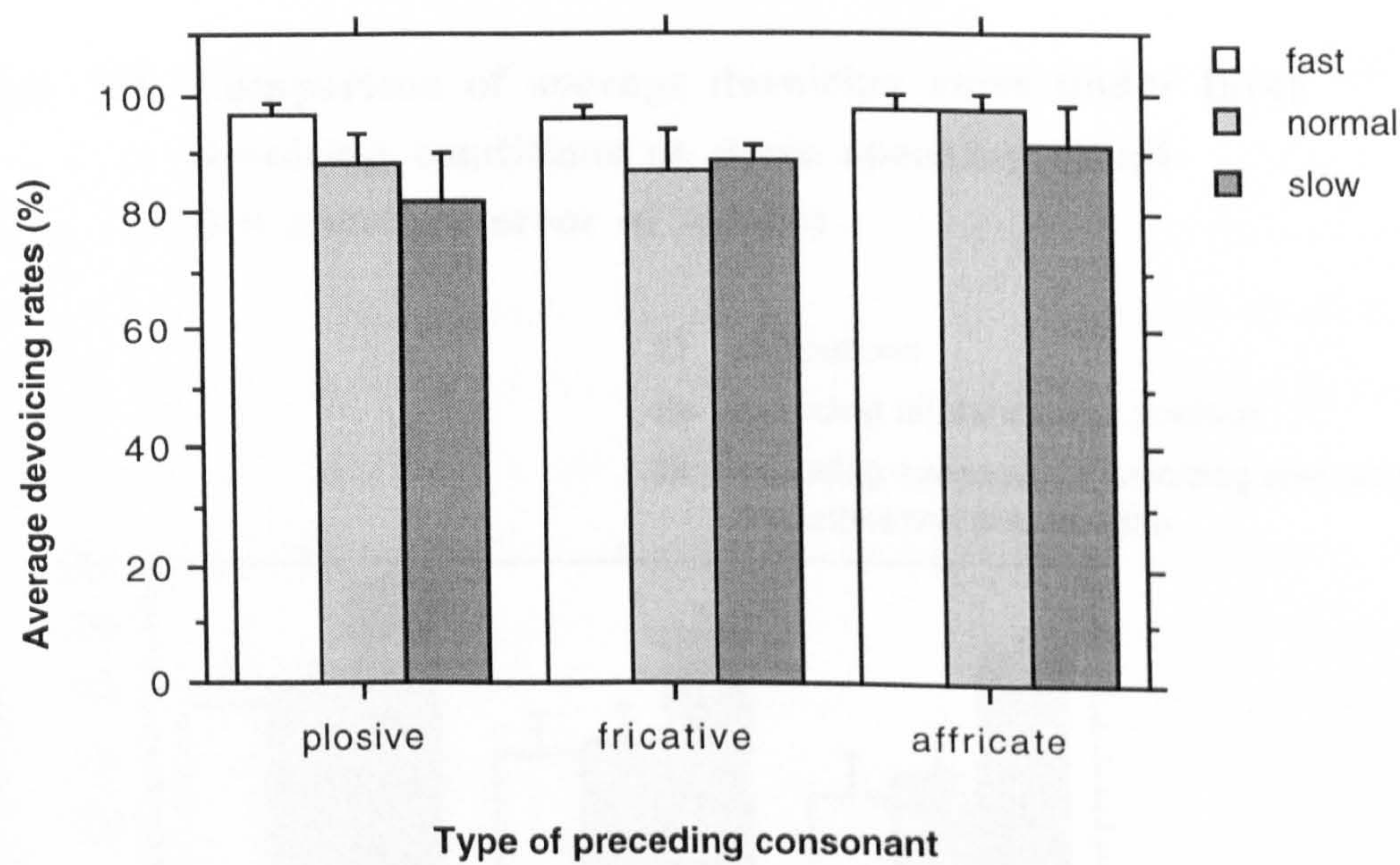
**Table 6.3 The total number of devoiced vowels of 6 subjects determined by the type of preceding consonant excluding utterance final positions and consecutive devoicing environments at three tempi**  
(Number of devoiced vowels/total number of examples)

<div>Tempo</div> <div>Type of consonant</div>	Fast	Normal	Slow
Plosive	153/159 examples (96.23%)	139/159 examples (87.42%)	128/157 examples (81.52%)
Fricative	168/177 examples (94.92%)	155/179 examples (86.59%)	144/164 examples (87.80%)
Affricate	104/107 examples (97.20%)	103/106 examples (97.17%)	86/97 examples (88.66%)



The results of a two-way ANOVA showed that there were no significant effects of speaking tempo [ $F(2,18)=2.223$ , n.s] nor consonant types [ $F(2,18)=.928$ , n.s] on devoicing rates. Also no significant effect was found in the interaction between speaking tempo and consonant types [ $F(4,18)=.406$ , n.s]. In other words, there was no effect of speaking rate on vowel devoicing regardless of the type of preceding consonants in the single devoicing condition. As discussed in Chapter 5, vowel devoicing seems to be an almost compulsory process as long as a devoiceable vowel is not in an utterance final position, and there are no devoiceable vowels in neighbouring morae, and the effect of speaking tempo is minimal.

**Figure 6.4** The devoicing rates determined by preceding consonant excludin consecutive devoicing environments and utterance final positions, at three tempi (I = Standard error of means)



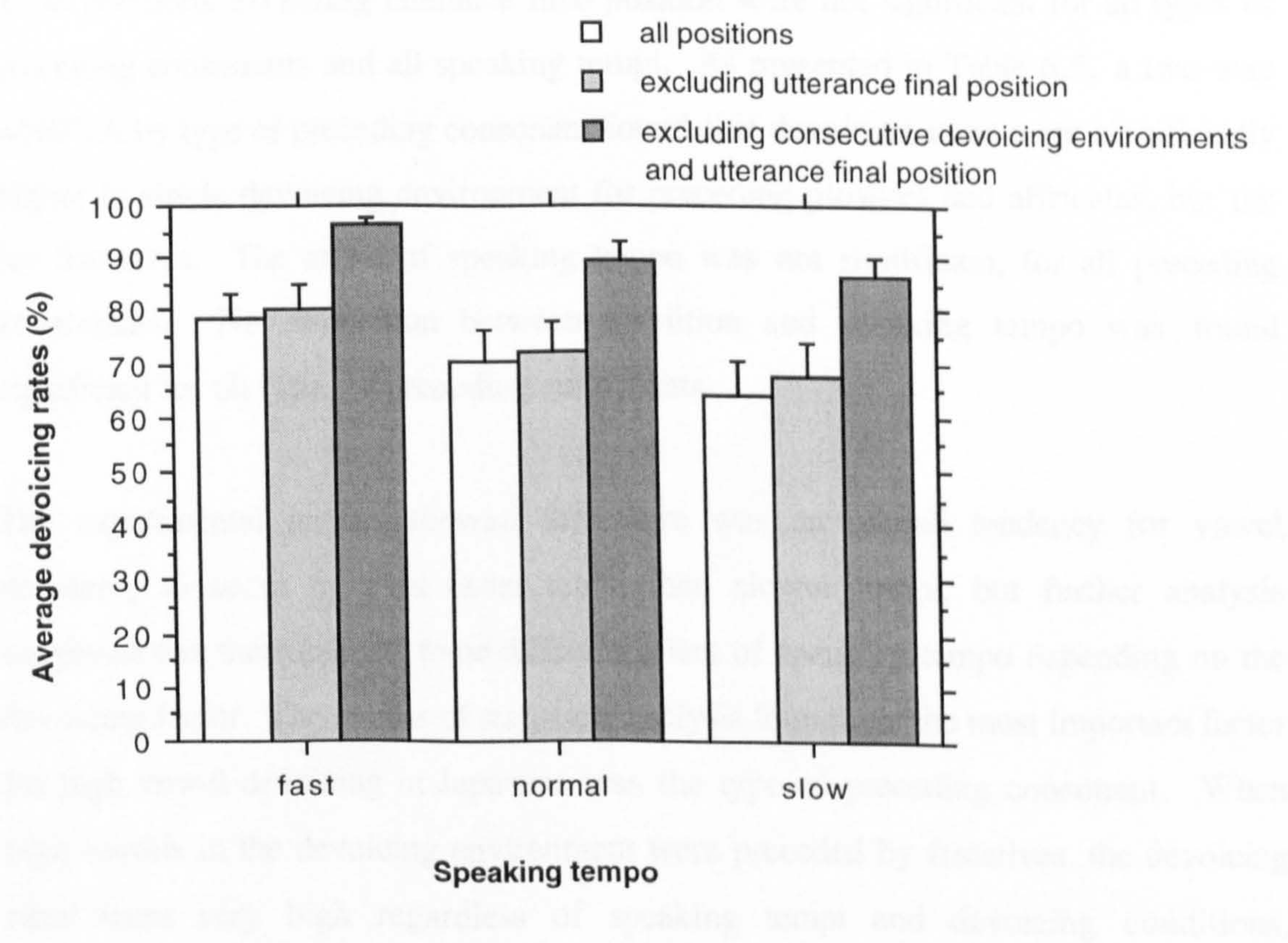
When the results were analysed by all three factors, namely (a) type of preceding consonants, (b) devoicing condition, and (c) speaking tempo, using a three-way ANOVA, the effect of all three factors were found to be significant. The devoicing rates were signigicantly higher when the vowel was preceded by fricatives, in the single devoicing environment and at the fast tempo. However, no interaction between the factors found significant. The statistical results are presented in Table 6.4.



Table 6.4 Three-way ANOVA table taking (a) type of preceding consonants, (b) devoicing condition, and (c) speaking tempo as factors

Factors	df	F-value	p-value
(a) Preceding consonant	2	15.893	p <.001
(b) Devoicing condition	2	22.594	p <.001
(c) Speaking tempo	2	5.574	p <.01
(a) x (b)	4	4.073	n.s.
(a) x (c)	4	.965	n.s.
(b) x (c)	4	.095	n.s.
(a) x (b) x (c)	8	.070	n.s.
Residual	66		

Figure 6.5 Comparison of average devoicing rates under three devoicing conditions at three speaking tempi (l = standard error of means)





When the results were analysed not considering the type of preceding consonants, as shown in Figure 6.5, the results of two-way ANOVA showed that devoicing rates were significantly higher at the fast tempo [ $F(2,84)=12.324$ ,  $p < .025$ ], and also significantly higher in the single devoicing condition [ $F(2,84)=4.217$ ,  $p < .001$ ]. However, no interaction between speaking tempo and devoicing condition was observed [ $F(4,84)=.059$ , n.s].

The devoicing rates excluding utterance-final positions and consecutive devoicing environments (Figure 6.4) were generally higher for all types of consonants than those including word-final positions and consecutive devoicing environments (Figures 6.1 and 6.2). However, when the effect of devoicing conditions by the type of preceding consonants was compared, a slight difference was observed.

Figures 6.6a, 6.6b and 6.6c show the average devoicing rates by the type of preceding consonant compared among three different devoicing conditions at three different speaking tempi. The differences of average devoicing rates between all positions and those positions excluding utterance final position were not significant for all types of preceding consonants and all speaking tempi. As presented in Table 6.5, a two-way ANOVA by type of preceding consonant found that devoicing rates were significantly higher in single devoicing environment for preceding plosives and affricates, but not for fricatives. The effect of speaking tempo was not significant, for all preceding consonants. No interaction between condition and speaking tempo was found significant for all types of preceding consonants.

The experimental results showed that there was an overall tendency for vowel devoicing to occur more at faster tempi than slower tempi, but further analysis suggested that there seemed to be different effect of speaking tempo depending on the devoicing factor. The results of statistical analysis found that the most important factor for high vowel devoicing in Japanese was the type of preceding consonant. When high vowels in the devoicing environment were preceded by fricatives, the devoicing rates were very high regardless of speaking tempi and devoicing conditions.

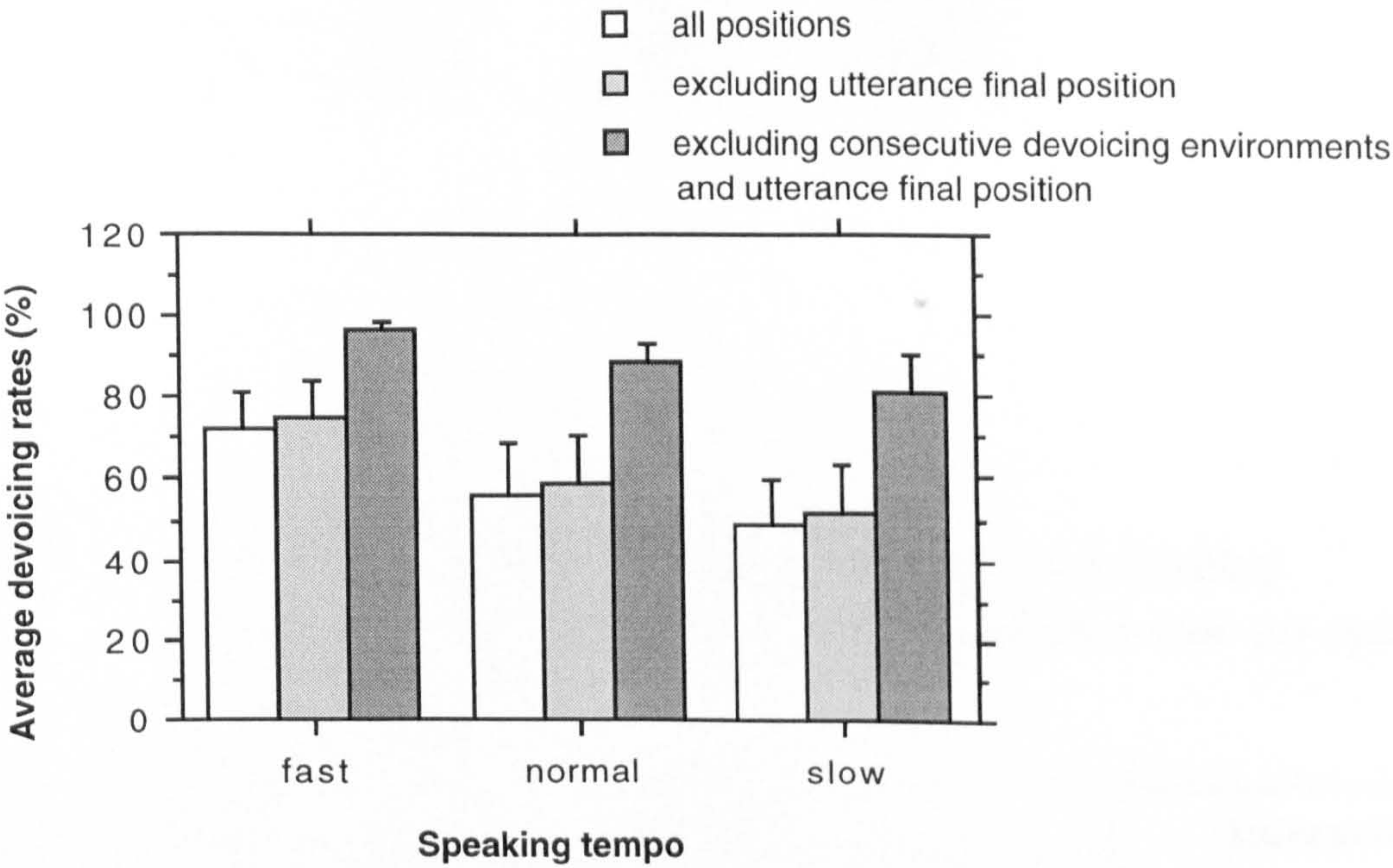


Therefore, devoicing rates were examined in the environment eliminating examples preceded by a fricative.

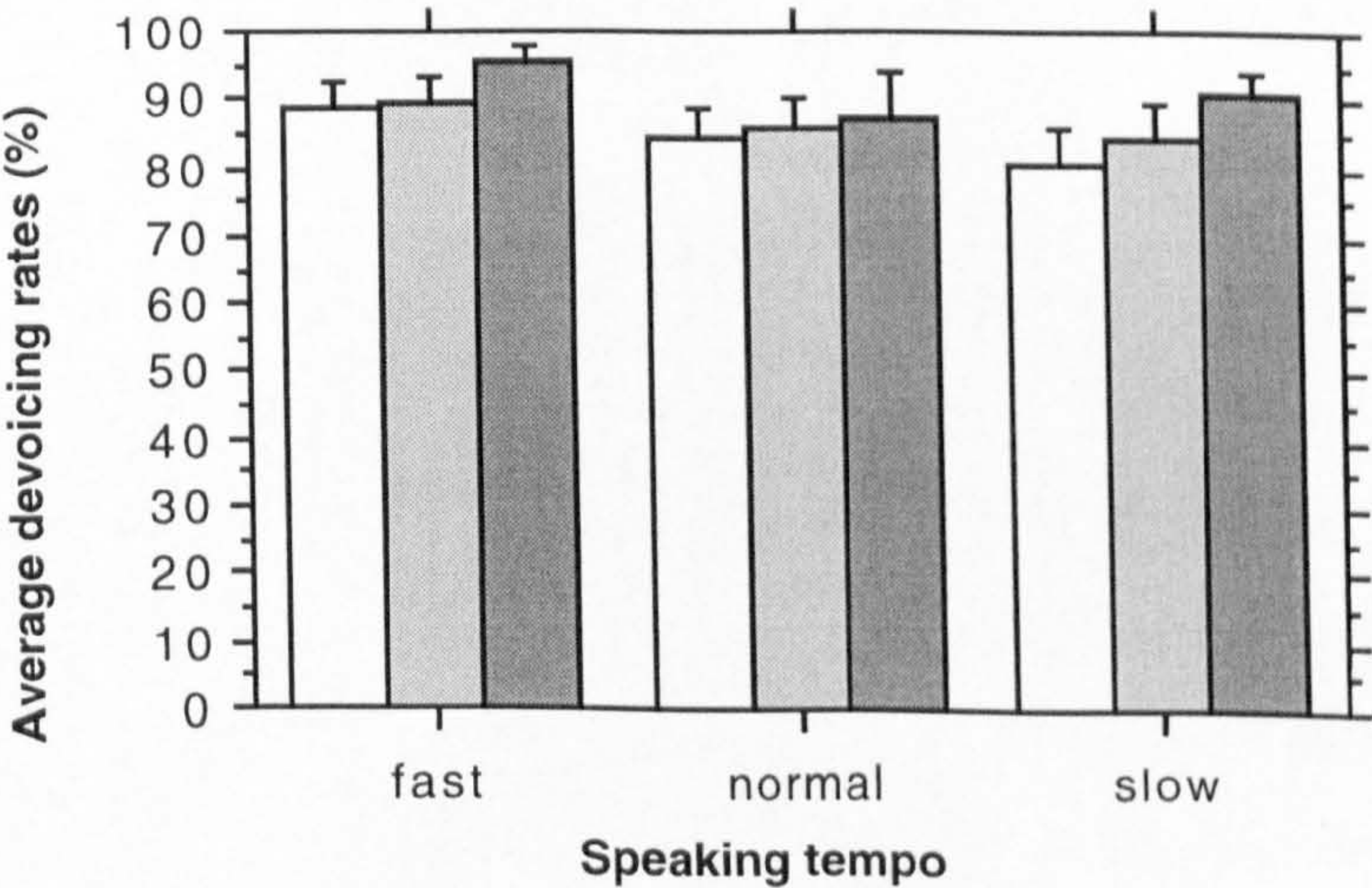
**Figures 6.6a, 6.6b and 6.6c**

**Comparison of the devoicing rates determined by the types of preceding consonants in three devoicing conditions, at three tempi (l = standard error of means)**

**(6.6a) Plosive**



**(6.6b) Fricative**



(Figure 6.6 continues)



(Figure 6.6 continued)

(6.6c) Affricate

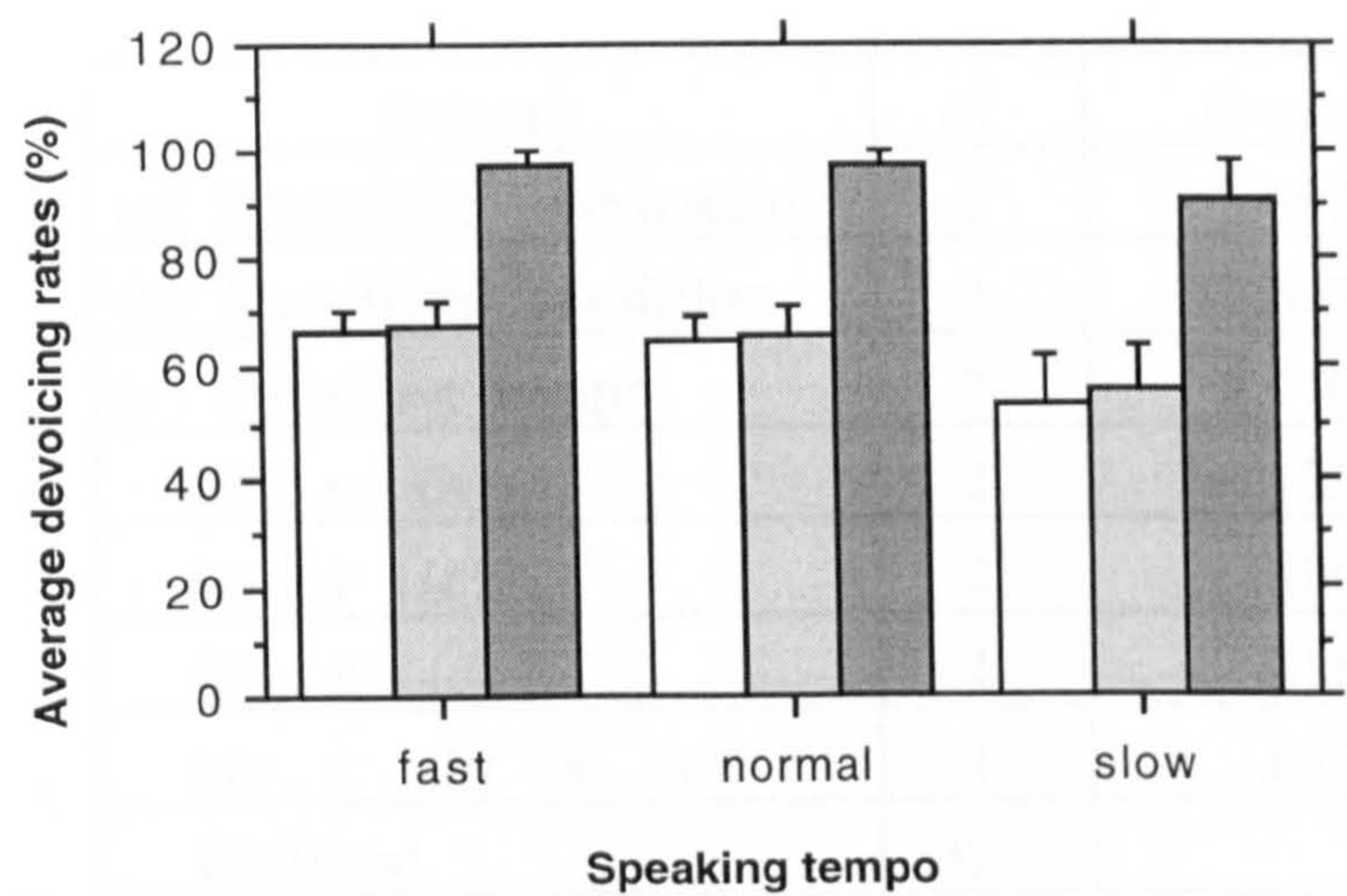


Table 6.5 Two-way ANOVA table by the type of preceding consonants taking (a) devoicing condition and (b) speaking tempo as factors

Preceding consonants	Factors		Interaction (a) x (b)
	(a) condition	(b) tempo	
Plosives	F(2,24)=7.167, p <.01	F(2,24)=3.107, n.s.	F(4,24)=.062, n.s.
Fricatives	F(2,33)=1.312, n.s.	F(2,33)=2.105, n.s.	F(4,33)=.116, n.s.
Affricates	F(2,9)=32.812, p <.001	F(2,9)=2.835, n.s.	F(4,9)=.125, n.s.

As shown in Table 6.6, three-way ANOVA showed that when the environment preceded by a fricative was eliminated, only the effect of devoicing condition was significant, but no significant effect was found in the type of preceding consonant, speaking tempo nor any interaction among the factors. It is interesting to compare this results with the results including examples preceded by fricatives (Table 6.4 and Figure 6.5), for which effect of tempo, devoicing condition, and type of preceding consonants were all significant.



**Table 6.6 Three-way ANOVA table excluding examples preceded by fricatives taking (a) type of preceding consonants, (b) devoicing condition, and (c) speaking tempo as factors**

<b>Factors</b>	<b>df</b>	<b>F-value</b>	<b>p-value</b>
<b>(a) Preceding consonant</b>	1	.087	n.s.
<b>(b) Devoicing condition</b>	2	12.699	$p < .001$
<b>(c) Speaking tempo</b>	2	2.710	n.s.
<b>(a) x (b)</b>	2	.135	n.s.
<b>(a) x (c)</b>	2	.406	n.s.
<b>(b) x (c)</b>	4	.088	n.s.
<b>(a) x (b) x (c)</b>	4	.019	n.s.
<b>Residual</b>	33		

The results by a three-way ANOVA (Table 6.6) suggested that the second important factor was the devoicing condition. The statistical results showed that devoicing rates were significantly higher when preceded by fricatives when all examples of devoiceable vowels were included (Figure 6.1) and also when utterance final positions were excluded (Figure 6.2), but the effect of preceding consonant type on vowel devoicing was not significant in the single devoicing condition (Figure 6.4). When examples preceded by fricatives were eliminated, the type of preceding consonants did not have an significant effect on vowel devoicing, but only devoicing condition was an significant factor (Table 6.6). This means that devoicing rates were normally significantly higher when devoiceable vowels were preceded by voiceless fricatives even in the condition excluding utterance final positions. However, in single devoicing conditions, there was no significant difference in devoicing rates regardless of type of preceding consonant; i.e. devoicing rates were significantly higher for all types of preceding consonants than when consecutive devoicing environments were included.

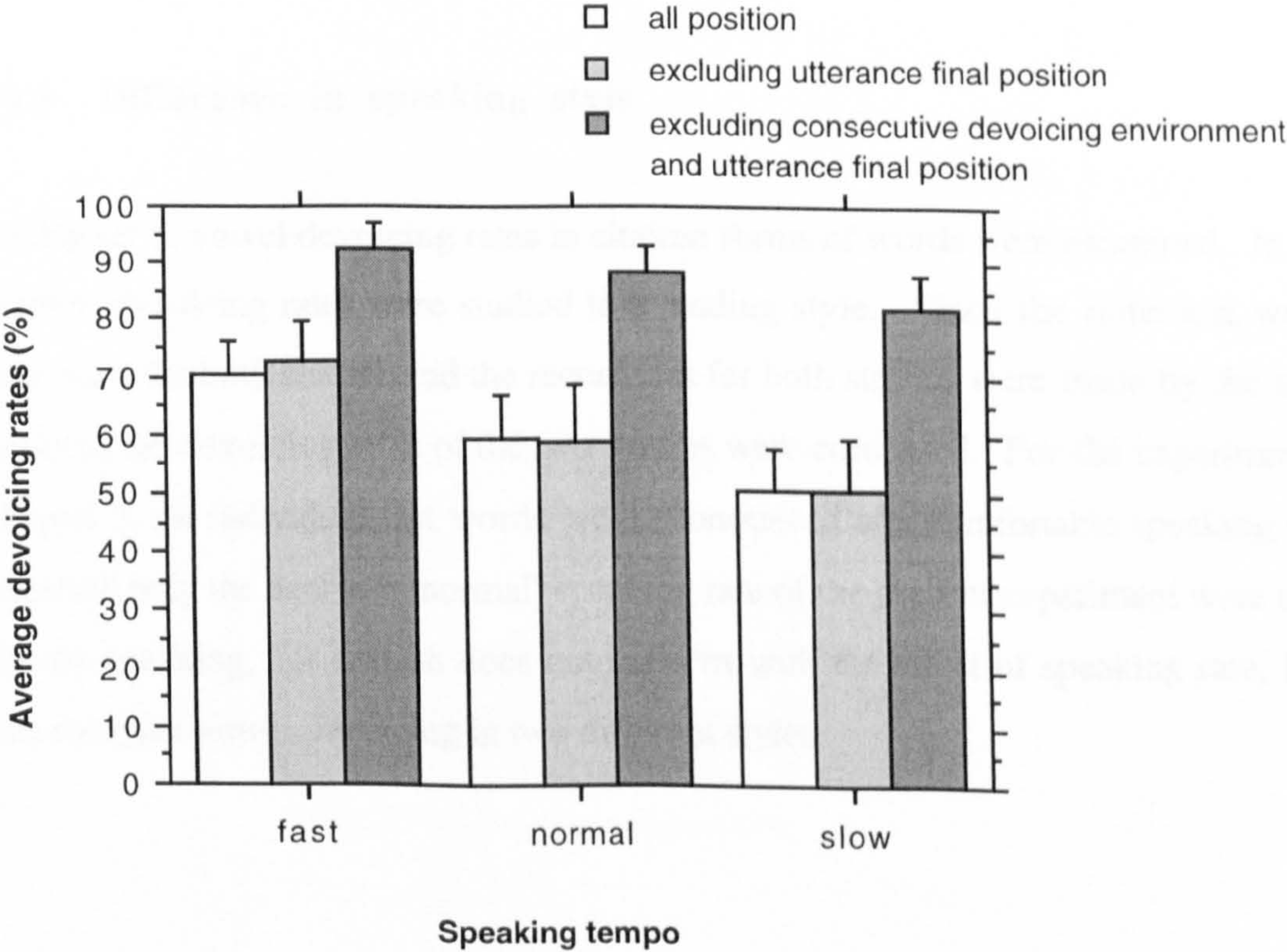
When the devoicing rates were analysed by speaking tempo and the devoicing condition excluding examples preceded by fricatives and without considering the type



of preceding consonants as shown in Figures 6.7 and 6.8, the effect of both the devoicing condition and speaking tempo were found to be significant in vowel devoicing. Devoicing rates were significantly higher in the single devoicing condition [ $F(2,42)=16.700$ ,  $p < .001$ ] and also at a fast tempo [ $F(2,24)=4.841$ ,  $p < .001$ ]. However, the interaction between devoicing conditions and speaking tempo was not significant [ $F(4,42)=.213$ , n.s.].

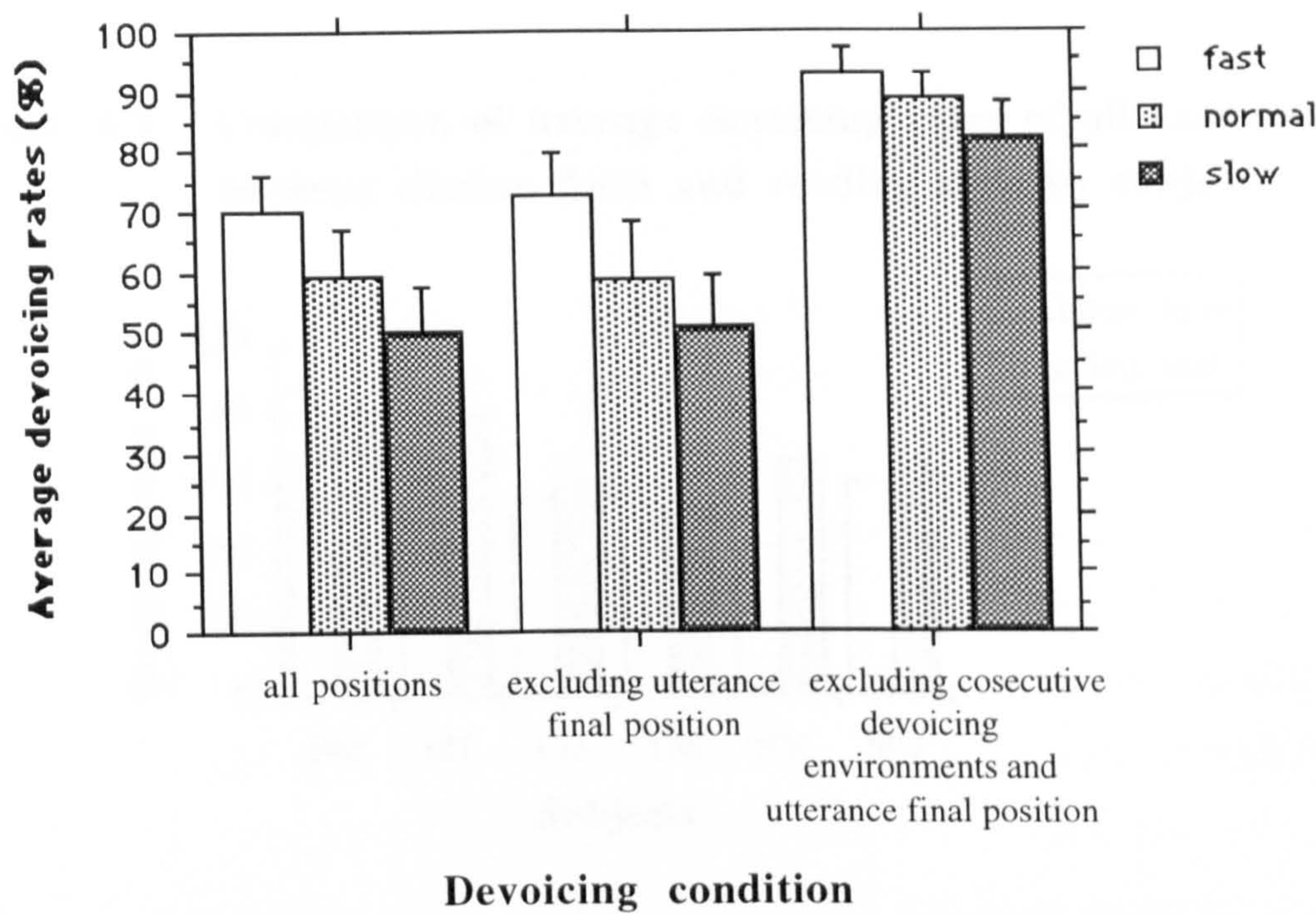
The experimental results suggested that generally average devoicing rates were highest at the fast tempo and lowest at the slow tempo, for all preceding consonant types and in all devoicing conditions, but statistical analysis did not find speaking tempo as an important factor in devoicing, except for the combined results of preceding plosives and affricates (see Figures 6.7 and 6.8). There were no significant interactions between tempo and other factors, namely type of preceding consonants and devoicing conditions.

**Figure 6.7** Average devoicing rates in environments preceded by plosives and affricates under three devoicing conditions at three tempi (I = Standard error of means)





**Figure 6.8** Average devoicing rates in cases preceded by plosives and affricates under three devoicing conditions at three tempi (I = Standard error of means)



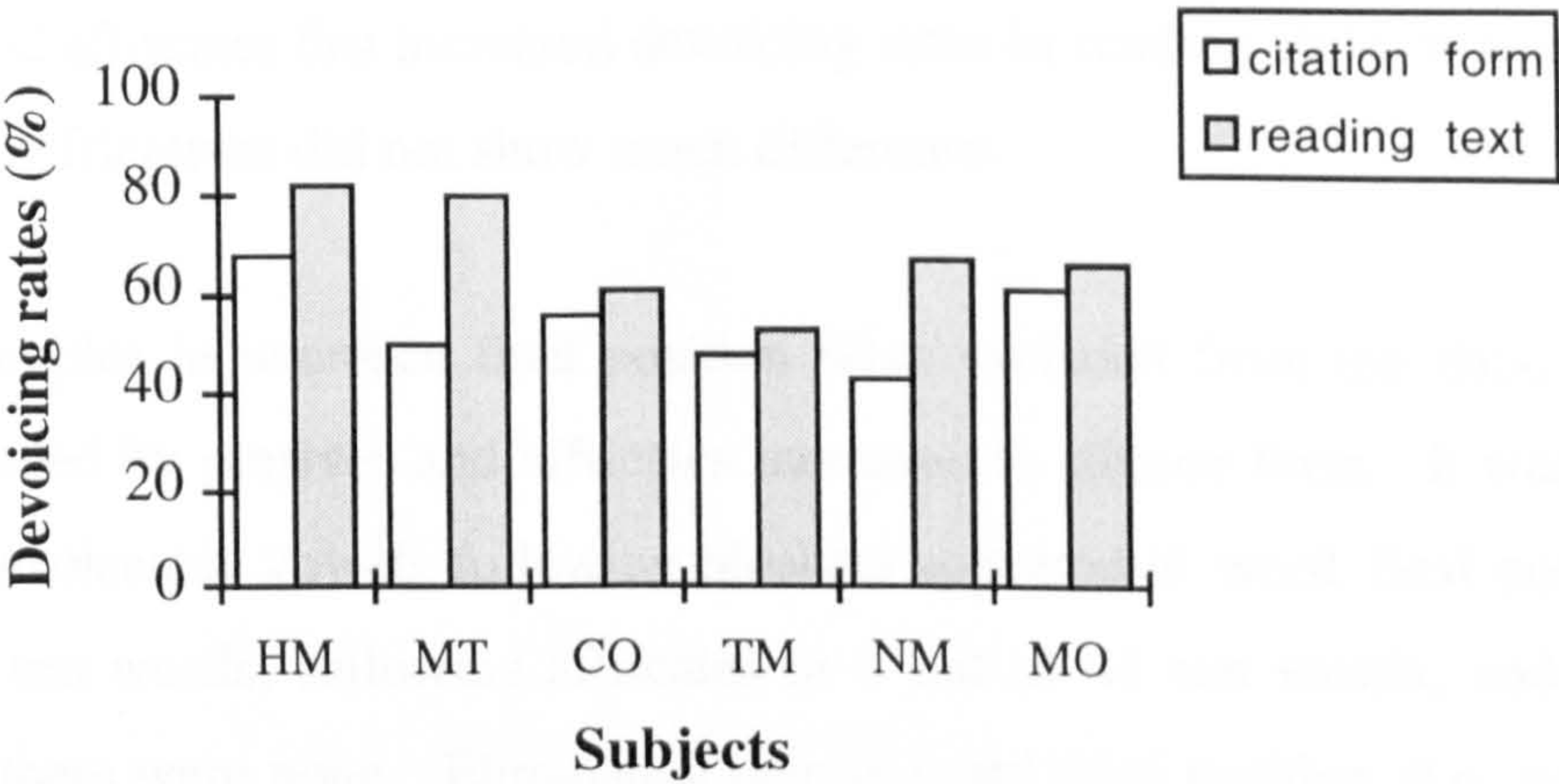
**6.4.4 Difference in speaking style**

In Chapter 5, vowel devoicing rates in citation forms of words were examined. In this chapter, devoicing rates were studied in a reading style. Since the same test words were used for both studies and the recordings for both studies were made by the same subjects, the devoicing rates of the two studies were compared. For the experiment in Chapter 5, the individual test words were pronounced at a comfortable speaking rate, therefore only the data at a ‘normal’ speaking rate of the present experiment were used. Strictly speaking, this section does not concern with the effect of speaking rate, but I will compare vowel devoicing in two different styles.

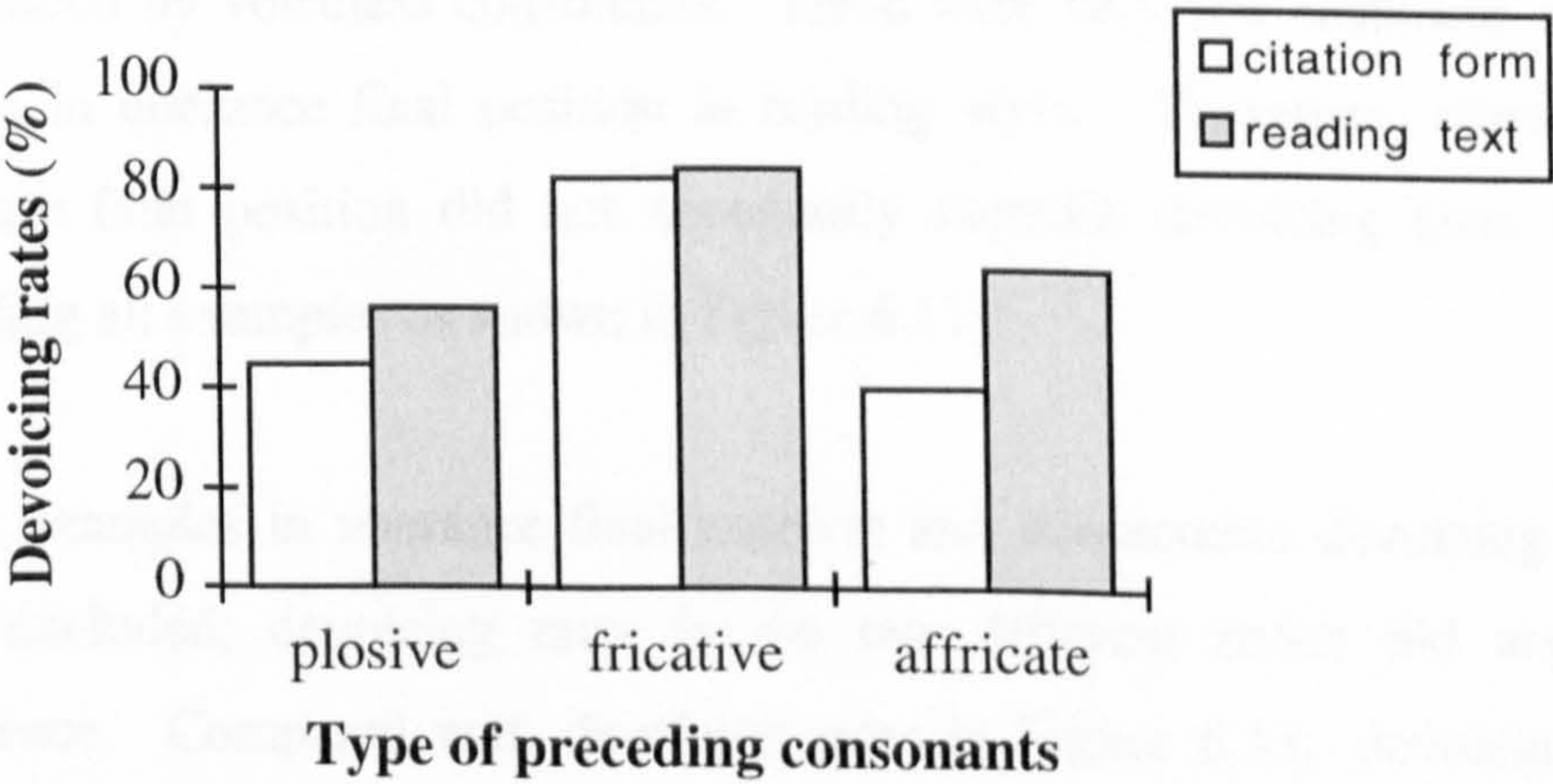


Figure 6.9 shows a comparison of devoicing rates between citation forms (Chapter 5, Figure 5.9) and reading style. All subjects devoiced more in reading style than in citation form, but the degree of increase varied from subject to subject. The increasing ratios varied from 7.9% for MO to 60% for MT.

**Figure 6.9      Comparison of average devoicing rates of all examples between citation form and reading style by subjects**



**Figure 6.10      Comparison of average devoicing rates of all examples of 6 subjects between citation form and reading style by the type of preceding consonants**



Style difference was compared by the type of preceding consonants and different devoicing conditions. Figure 6.10 shows average devoicing rates of 6 subjects by the type of preceding consonants. Devoicing rates were generally higher in reading style than in pronouncing citation form of words, but very little difference was observed in cases preceded by fricatives. When devoicing rates were studied within the same speaking style, preceding fricatives triggered devoicing significantly more than plosives and affricates. As shown in Figure 6.9, all subjects increased the devoicing rates in reading style from that in citation form, and it was vowels preceded by plosives and affricates that increased devoicing rates in reading style, whereas vowels preceded by fricatives did not show much difference.

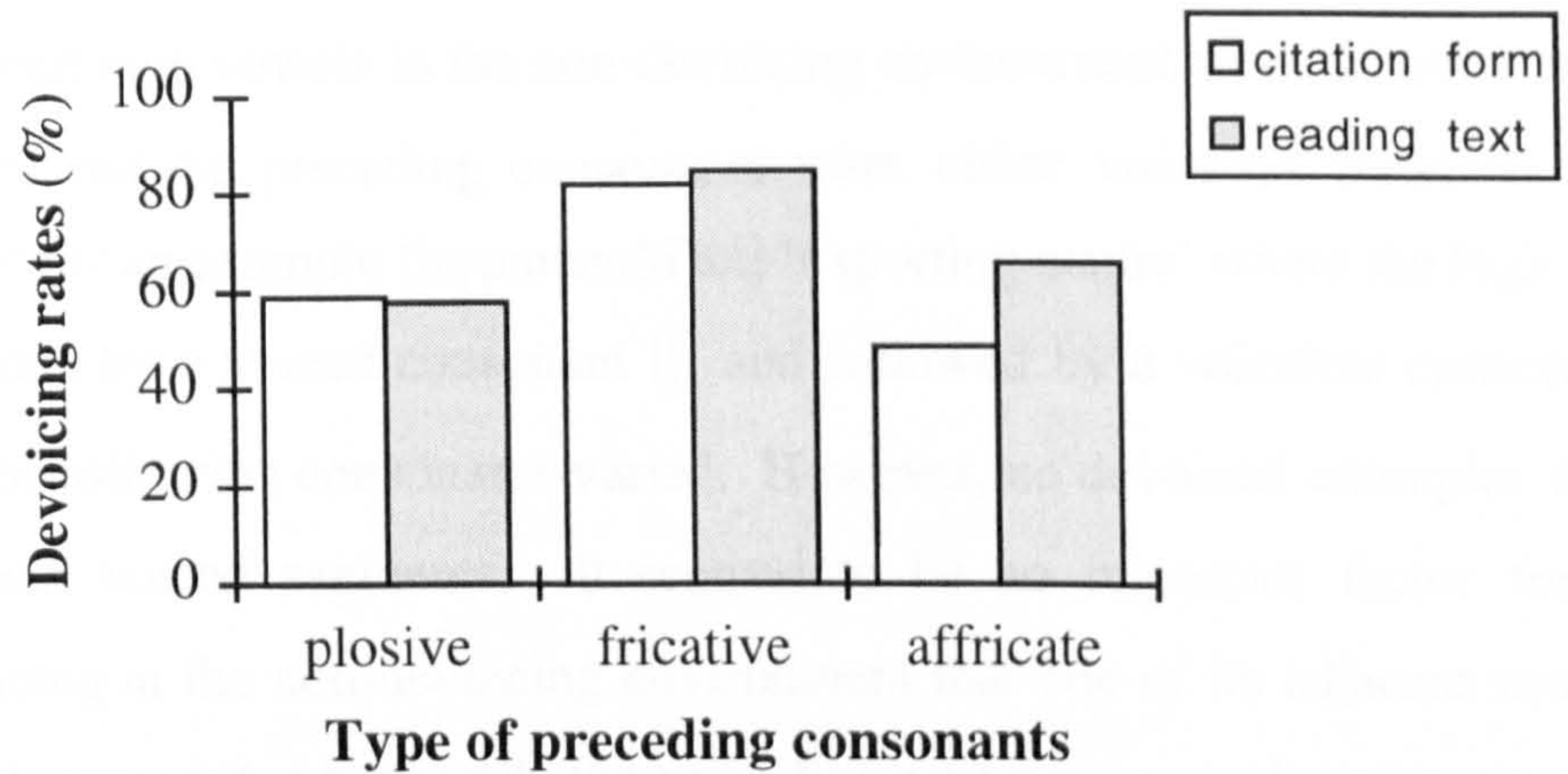
When examples in utterance final position were excluded from the data, devoicing rates preceded by plosives and affricates increased in citation form. It was probably because voiceable vowels following plosives appeared in word final position in 8 out of 41 test words, following affricates in 4 out of 41 test words, and following fricatives there were none. Eliminating data in word final position (i.e. an utterance final position in citation form, where devoicing was less likely to occur) may have increased devoicing rates of the vowels following plosives and affricates. On the other hand, in the reading style, word final positions were not necessarily utterance final positions. Most of test words, except for a word /kiri hukiki/ 'sprayer', were followed by other words, some of which started with voiceless consonants, which created an ideal devoicing environment for word final vowels, i.e. utterance internal position surrounded by voiceless consonants. There were very few instances of voiceable vowels in utterance final position in reading style. Therefore, eliminating data in utterance final position did not necessarily increase devoicing rates from the data including all examples as shown in Figure 6.11.

When examples in utterance final position and consecutive devoicing environments were excluded, devoicing rates in the two different styles did not show much difference. Compared with devoicing rates in Figure 6.11, devoicing rates in the single devoicing environment in Figure 6.12 were much higher regardless of type of



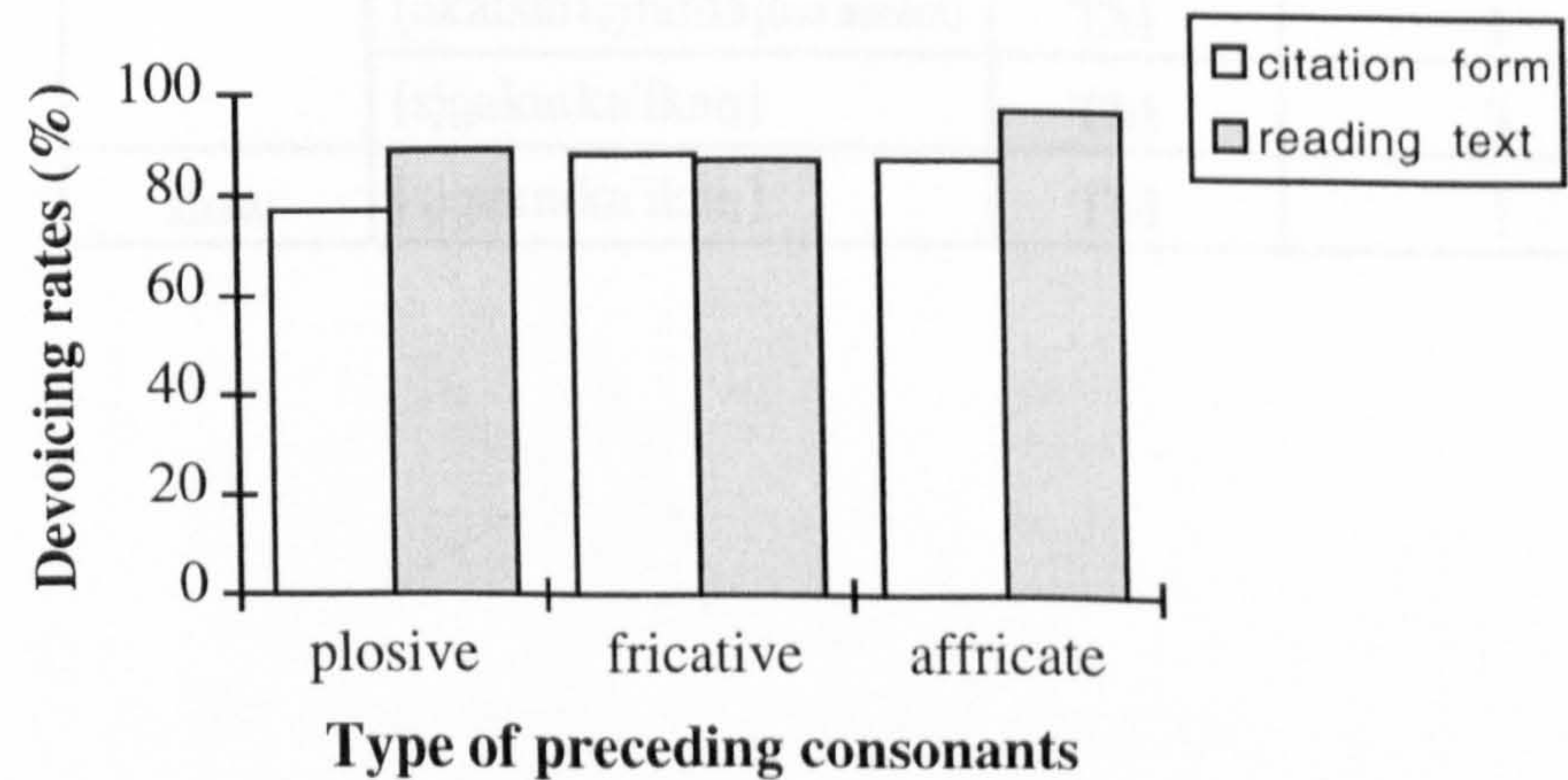
preceding consonants and speaking style. The effect of speaking style seems to be minimal in the single devoicing condition.

**Figure 6.11** Comparison of average devoicing rates of all examples of 6 subjects between citation form and reading style by the type of preceding consonants excluding utterance final position



**Table 6.12** Voicing of the preceding consonant

**Figure 6.12** Comparison of average devoicing rates of all examples of 6 subjects between citation form and reading style by the type of preceding consonants excluding utterance final position and consecutive devoicing environments





#### 6.4.5 High vowel devoicing in non-devoicing environments

There were quite a few instances of high vowels being completely voiceless or almost voiceless in non-devoicing environments. This occurred in five speakers at all speaking rates for both /i/ and /u/ and is listed below in Tables 6.7a and 6.7b.

Although high vowel devoicing in non-devoicing environments occurred at all tempi, it was more frequent at the faster tempi than the slower tempi. Typical environments for devoiced high vowels in the non-devoicing environment were in word initial and final morae, and the preceding consonants were either voiceless fricatives or affricates, except for an example [jɯɕʷʊtsʊkiʔtɕi] 'exporting centre' where the high vowel /u/ was preceded by a voiced consonant [j] and followed by a voiceless consonant [ɕ]. The type of following consonants varied. However, no devoiced examples were observed between voiced segments. It seemed to be an important factor for high vowel devoicing in the non-devoicing environment that one of its adjacent sounds has to be voiceless, and that preferably the preceding sound is a voiceless fricative or affricate.

**Table 6.7 (a) Voiceless /i/ in non-devoicing environments**

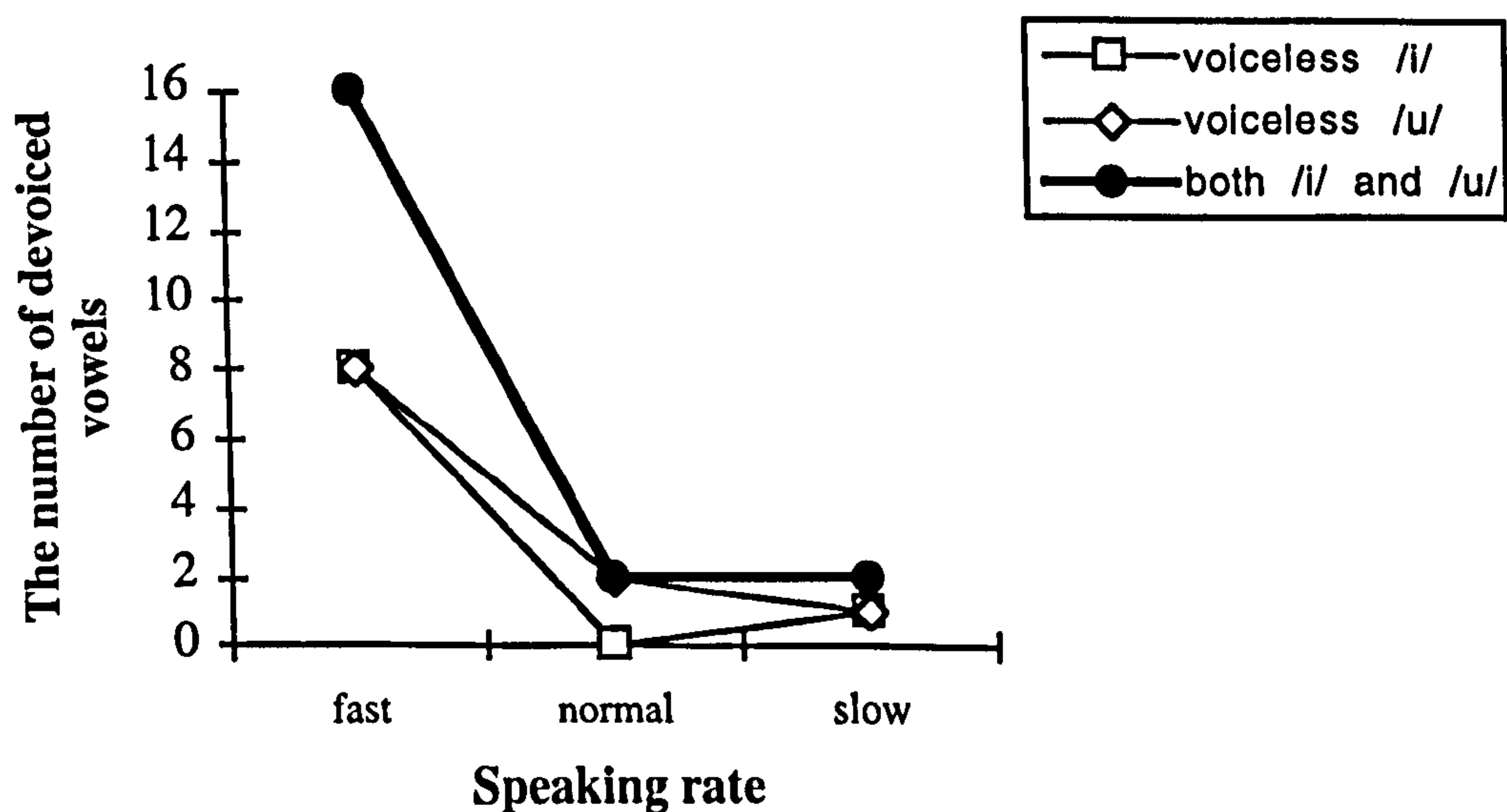
tempo	test word	subject	No. of occurrence
fast	[kɿriɸʊkiʔki]	HM	1
		TM	3
	[akatsʊtɕɿjama](no accent)	TM	1
	[sɿgakʊkaʔikaŋ]	TM	3
slow	[sɿgakʊkaʔikaŋ]	TM	1

**Table 6.7 (b) Voiceless /u/ in non-devoicing environments**

tempo	test word	subject	No. of occurrences
fast	[suɰizoˈkʷaŋ]	TM	3
	[ɕo:ɕikeˈ:kakɰ (ni)]	HM	3
	[tsʉkʉɕiteˈtsʉdo:]	MO	2
normal	[jʉɕʉʉtsʉkiˈtɕi]	MO	1
	[do:kʉʉtsʉ(no)](no accent)	MT	1
slow	[jʉɕʉʉtsʉkiˈtɕi]	NM	1

The effect of speaking rate on high vowel devoicing in the non-devoicing environment was examined using a Chi-square test. The total number of examples were calculated as: “the number of test words with devoiceable high vowels in the non-devoicing environment (8)” x “3 repetitions” x “6 subjects” = 144, except for the case that Subject TM mispronounced a test word /sigakukaikaN/ in one instance at the slow tempo. Table 6.8 shows observed and expected numbers of voiced and devoiced high vowels in non-devoicing environments at three tempi in two conditions (devoiced and voiced).

**Figure 6.13 The number of devoiced high vowels in non-devoicing environments at three different tempi**



**Table 6.8 Observed and expected data of devoiced and voiced high vowels /i/ and /u/ in the non-devoicing environment at three different tempi (the numbers in square brackets are cell numbers)**

	Devoiced vowels	Voiced vowels	Total
Fast (expected data)	16 [1] (6.68)	128 [2] (137.32)	144
Normal (expected data)	2 [3] (6.68)	142 [4] (137.32)	144
Slow (expected data)	2 [5] (6.64)	141 [6] (136.36)	143
Total	20	411	431

The results showed that there was a significant difference in the numbers of devoiced and voiced high vowels in the non-devoicing environment depending on the speaking rate ( $\chi^2=20.551$ ,  $df=2$ ,  $p <.001$ ). The value for each cell was as (1):

$$\begin{aligned} (1) \chi^2 = & 12.993 \text{ (Cell 1)} + 0.632 \text{ (Cell 2)} + \\ & 3.281 \text{ (Cell 3)} + 0.160 \text{ (Cell 4)} + \\ & 3.239 \text{ (Cell 5)} + 0.158 \text{ (Cell 6)} = 20.462 \end{aligned}$$

The values of the cells implied that the number of devoiced high vowels in the non-devoicing environment at the fast tempo occurred significantly more than those at the normal and slow tempi, but no difference was observed at the normal and slow tempi.

**6.4.6 Devoicing of non-high vowels**

Devoicing of non-high vowels /a/ and /o/ was observed at all tempi for all speakers, and is listed in Tables 6.9a and 6.9b.



**Tables 6.9 The list of devoiced non-high vowels in devoicing environments**

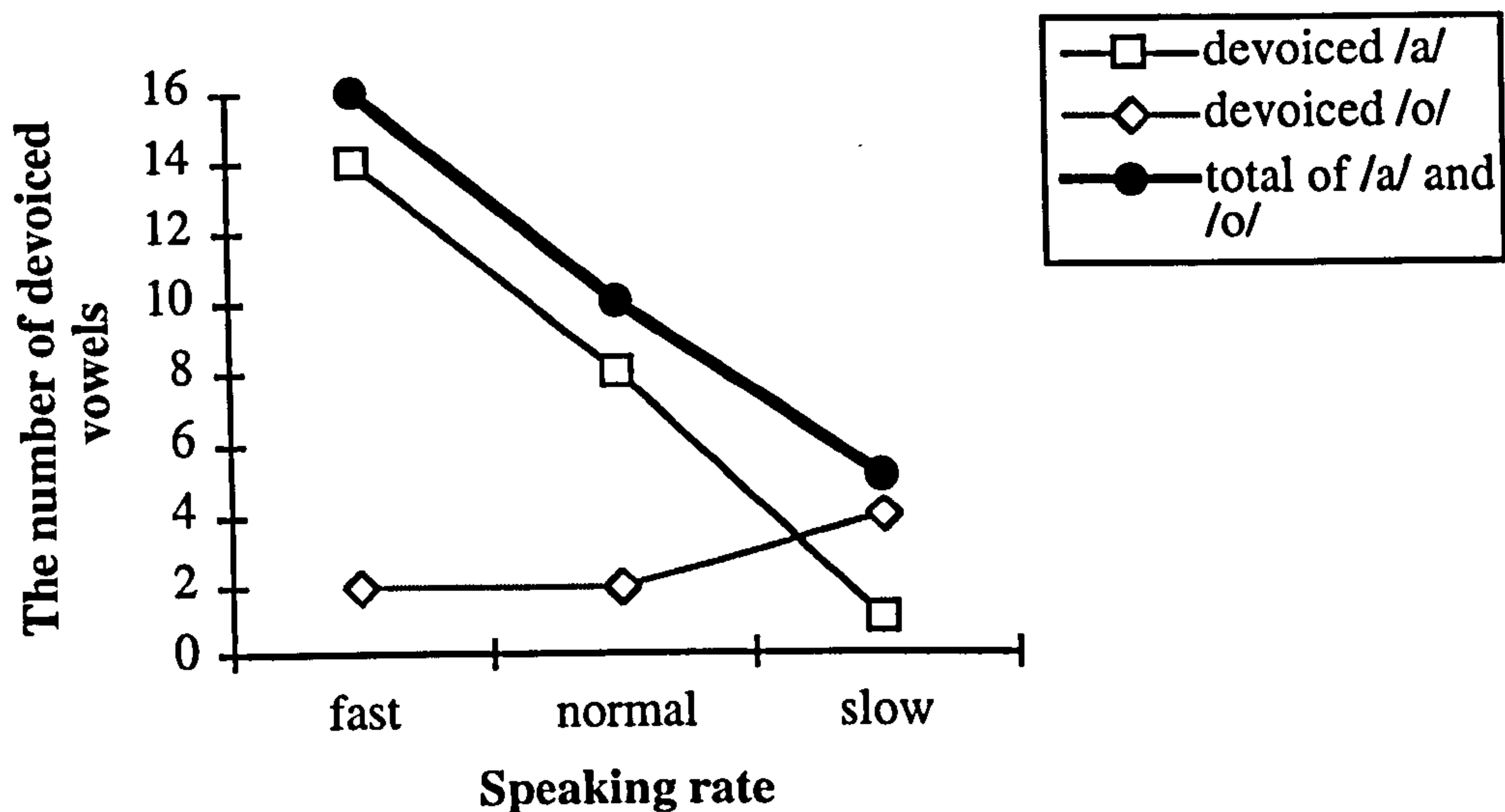
**(6.9a) Completely or almost voiceless [q]**

tempo	test word	subject	No.of occurrences
fast	[çqki'kaŋ]	TM	1
	[to:hokwɪtçi'hq:]	HM	1
normal	[çqki'kaŋ]	HM	1
	[kqkwɸwkw] (no accent)	CO	1
slow	[kqkwɸwkw] (no accent)	TM	1
		CO	2
	[natçi'ke':kqkw]	CO	1

**(6.9b) Completely or almost voiceless [q̥]**

tempo	test word	subject	No.of occurrences
fast	[t̥akakwke':e:]	HM	3
		TM	2
		NM	1
	[k̥akwjaswɪtsu'a:]	HM	1
		TM	1
	[k̥aswɪpi'kai]	HM	1
		MO	1
	[̥akatswɪtçi'jama](no accent)	HM	3
	[ço:çi'ke':k̥akw]	HM	1
normal	[t̥akakwke':e:]	HM	3
		TM	3
		CO	1
	[k̥akwjaswɪtsu'a:]	NM	1
slow	[k̥aswɪpi'kai]	MO	1

**Figure 6.14** The number of devoiced non-high vowels [q] and [ɤ] in devoicing environments at three different tempi



As far as the vowel /a/ is concerned, devoicing was more frequent at faster tempi than at slower tempi. Devoicing of non-high vowels occurred only in certain environments, i.e. in a word-initial unaccented mora and/or between voiceless consonants. When a non-high vowel was devoiced in a word-internal position, the vowel was between the same voiceless consonant phonemes, such as [çɔ:çike':kɤkɯ] and [natçike':kɤkɯ]. The only exception was in [to:hokɯtçi'hɔ:] where the final /o/ was a long vowel and in a word-final position. In this particular utterance, the phonation of the subject HM was gradually changed from modal to whisper, so half of the word was whispered rather than the last vowel /o:/ being devoiced<sup>2</sup>.

Table 6.10 presents observed and expected numbers of voiced and devoiced non-high vowels at three tempi in two conditions (devoiced and voiced vowels). The results showed that there was a significant difference in the numbers of devoiced and voiced high vowels in the non-devoicing environment depending on the speaking rate ( $\chi^2=6.271$ ,  $df=2$ ,  $p < .05$ ).

<sup>2</sup> For the definitions of devoiced and whispered speech, see Chapter 1.

**Table 6.10**    Observed and expected data of devoiced and voiced non-high vowels /o/ and /a/ in the devoicing environment at three different tempi (the numbers in square brackets) are the cell numbers)

	Devoiced vowels	Voiced vowels	Total
<b>Fast</b> (expected data)	16 [1] (10.33)	146 [2] (151.67)	<b>162</b>
<b>Normal</b> (expected data)	10 [3] (10.33)	152 [4] (151.67)	<b>162</b>
<b>Slow</b> (expected data)	5 [5] (10.33)	157 [6] (151.67)	<b>162</b>
<b>Total</b>	<b>31</b>	<b>455</b>	<b>486</b>

The value for each cell was as (2):

$$\begin{aligned} (2) \chi^2 &= 3.108 \text{ (Cell 1)} + 0.212 \text{ (Cell 2)} + \\ &\quad 0.011 \text{ (Cell 3)} + 0.001 \text{ (Cell 4)} + \\ &\quad 2.753 \text{ (Cell 5)} + 0.188 \text{ (Cell 6)} = 6.271 \end{aligned}$$

The values of the cells implied that there was a tendency for the number of devoiced non-high vowels in the devoicing environment at the fast tempo to occur more frequently than those at the normal and slow tempi, and less frequently at the slow tempo than those at the normal and fast tempi.

Non-high vowel devoicing was not as consistent as high vowel devoicing and was limited to certain phonetic environments, i.e. phonetically weak position such as an unaccented word-initial or a phonetically ideal environment for devoicing such as between voiceless consonants, whereas high vowel devoicing occurred in various positions in a word and even when adjacent to a voiced consonant (see Chapter 5). The analyses of spectrograms and speech pressure waveforms in the earlier experiment in Chapter 5 showed that the amplitude of sounds generally tends to be weaker in word-initial and word-final positions. Everything being equal, non-high vowels are universally intrinsically longer than high vowels. Therefore, non-high vowels



probably require more phonetically desirable environments to be devoiced. It seems that non-high vowel devoicing is a result of glottal assimilation among neighbouring sounds rather than a phonological process. Therefore, it seems plausible for the non-high vowel devoicing to be more frequent at faster tempi than slower tempi.

#### 6.4.7 Problems of controlling speaking rate

For this experiment, the subjects were asked to control speaking rates by themselves because a comfortable speaking rate varies from speaker to speaker. As a result, there were some problems in controlling tempi. First of all, it seemed to be difficult for all subjects to keep a constant speaking tempo, especially at fast and slow tempi. They started off at fast and slow tempi, but tended gradually to revert to normal tempo. It was noticeable that especially after a subject made a mistake, he tended to slow down a little. Generally, the total duration of the texts was shortest at the fast tempo and longest at the slow tempo including repetition time caused by making mistakes for all speakers. Considering the fact that the subjects made more mistakes at the fast tempo than slower tempo, i.e. more repetitions at faster tempi, they controlled speaking tempi reasonably well.

Secondly, there was a tendency for all subjects to change the duration of pauses, rather than the actual speaking rate. At a fast tempo, the subjects did not pause very long; they tended to read sentences one after another without taking enough of a pause. On the other hand, at a slow tempo, not only did they take a long pause between sentences, but also they inserted a pause more often between phrases, words and at internal word boundaries. Therefore, the actual speaking rate of each word or sentence may not have differed very much among the three tempi. However, this may be the reality of speech at difference tempi: fast speech actually means taking a shorter pause and slow speech actually means taking a longer pause.

Moreover, the speakers did not insert pauses wherever they liked. They tended to slow down and pause between sentences, phrases, words, and at internal word

boundaries. When they inserted a pause between words and at internal word boundaries, this made the preceding mora utterance final, where devoicing rates of high vowels tend to be lower (see Chapter 5). When a subject inserted a pause in a given position in one utterance but did not in another, the devoiceable vowel in the preceding mora was usually voiced in a prepausal position (utterance final), but usually voiceless in the other utterance (non-utterance final). As suggested in the hypothesis (4), some blocking factors for Japanese vowel devoicing, such as position in a word, are more sensitive to speaking tempo than others (see Section 6.2). A slow tempo may have directly caused the lower devoicing rates, but it may also have created an ideal environment in which to block vowel devoicing such as pausing at an internal word boundary, and indirectly lower devoicing rates in devoicing environments.

#### **6.4.8 Is high vowel devoicing an optional process of fast speech?**

The results of this experiment suggested that vowel devoicing generally occurred more at faster tempi than slower tempi. However, the statistical results showed there was no significant effect of speaking rate on overall devoicing rates, regardless of the types of preceding consonant.

Further analysis suggested that devoicing rates were also influenced by other devoicing factors, i.e. position in an utterance and devoicing conditions. The statistical results showed that the effect of speaking rates was not generally significant except for the effect of devoicing condition on preceding plosives and affricates (see Figures 6.6a, 6.6c and 6.8), either when the utterance final position was excluded, or when both the utterance final position and consecutive devoicing environments were excluded. When consecutive devoicing environments were excluded, i. e. in the single devoicing condition, there was no significant effect of preceding consonant type, nor speaking rate (see Figure 6.4). As far as the single devoicing condition is concerned, high vowel devoicing seems to be a virtually compulsory process at all tempi; not an optional fast speech rule.

It was under consecutive devoicing conditions that there were differences in devoicing rates at different speaking tempi. When devoiceable vowels were preceded, not by fricatives, but by plosives or affricates in the consecutive devoicing condition, devoicing rates were significantly higher at faster tempi and lower at slower tempi; in other words, vowel devoicing seems to be a fast speech rule in the consecutive condition. This result leads to two questions: (1) why is there no effect of speaking tempo when a preceding consonant is a fricative?, and (2) why is there a different effect of speaking tempo depending on devoicing conditions?

The previous experiment in Chapter 5 showed that in consecutive devoicing environments, not all devoiceable vowels became voiceless in the reading of citation forms of words at normal speaking rate. Blocking factors of vowel devoicing such as the presence of an internal word boundary, accentedness and position in an utterance become effective and blocked some devoiceable vowels from undergoing the devoicing process. In the consecutive devoicing environment, vowel devoicing could occur consecutively over one mora but devoicing of more than two consecutive morae was not favoured (see Chapter 5, Section 5.4). The results of this experiment also found different effects of devoicing factors among different devoicing environments, such as the type of preceding consonant, devoicing conditions, and speaking rates.

As far as type of preceding consonant is concerned, a preceding fricative seems to trigger devoicing of a following devoiceable vowel in all conditions, at all tempi. The results agree with studies such as Takeda and Kuwabara (1987) and Yoshida and Sagisaka (1990), as well as the result of the previous experiment.

Physiological studies on Japanese vowel devoicing demonstrated that vowel devoicing involved slightly different laryngeal muscular activity, depending on the type of preceding consonant. For instance, Yoshioka et al. (1982) studied the glottal opening gesture during the sequences of voiceless sounds, /kikee/, /kisee/, /sikee/, /sisee/ in which the high vowel /i/ was devoiced in a carrier sentence, using the combined techniques of photo-electric glottography, fiberoptic filming and laryngeal electromyography. They measured glottal opening width and glottal opening velocity.



Results by Yoshioka et al. found that the peak activity of the posterior cricoarytenoid muscle (the abductor muscle) occurred at the voicing offset of the last vowel of the carrier sentence, i.e. the vowel preceding to the initial consonant of the test words, regardless of type of initial consonants. However, their results showed the difference in the peak value of the opening gesture and the speed of the glottal movement depending on the type of consonants surrounding the devoiced vowel. When a voiceless fricative preceded the devoiced vowel, the glottal opening movement was faster than when preceded by a voiceless plosive. Moreover, the peak value of the opening gesture was larger when the devoiceable vowel was surrounded on both sides by voiceless fricatives than plosives, and the timing of the peak opening was early and relatively fixed for the test words beginning with a fricative but the timing was comparatively late and more variable for the test words beginning with a plosive. As for the speed of the glottal movement, the velocity during the opening phase was faster for the test words beginning with a fricative than for the test words beginning with a plosive.

Yoshioka et al. (1982: 9) concluded that “the faster and larger opening for a voiceless fricative may be related to the necessary supply of air during the voiceless fricative segment to produce adequate turbulent noise by a quick reduction of laryngeal resistance. On the other hand, in order to stop glottal vibrations at the implosion of a voiceless stop and to assist in the buildup of oral pressure, a slight opening gesture may be sufficient in combination with the closing gesture of the supraglottal articulators”.

Moreover, Kuehn and Moll (1976) studied the effect of phonetic context on the velocity of movements of the tongue, lips and mandible in a CV transition and a VC transition, and found that as far as the transitional velocities in /CV/ and /VC/ are concerned, articulator movements into and out of a fricative constriction were usually slower than those of stops since fricatives are characterised by vocal tract constrictions of carefully controlled and relatively constant size. Based on their findings, Beckman proposed:

When an oral constriction that is narrow enough to give rise to frication is slowly released into the only somewhat less constricted oral configuration for a short high vowel, the oral air pressure that has built up behind the constriction cannot be vented as rapidly as it can than when the consonant is a stop or when the release is into a more open vowel. Well into the vowel, aerodynamic conditions may remain more conducive to continued turbulence at the oral constriction than to vocal fold vibration, and if the vowel is short enough, it can be effectively devoiced or deleted (Beckman, 1993: 13)

The results of faster and larger movements of the glottal muscles for preceding fricatives by Yoshioka et al. provide an ideal glottal state for devoicing and slower movements of the tongue, lips and mandible for preceding fricatives by Kuehn and Moll assist further ideal environment for vowel devoicing than preceding plosives.

Weitzman et al. (1976) studied glottal opening width during the production of devoiced vowels in normal and whispered speech in Japanese during the sequence of /CVtee/ (where C is a voiceless consonant, and V is a devoiceable vowel) in words such as such as /kitee/, /sitee/ and /titee/<sup>3</sup>. The abduction reached its maximum peak during the [ÇV] interval in both normal and whispered speech. When /C/ was a plosive, the maximum glottal opening usually occurred during the devoiced vowel interval, whereas when /C/ was a fricative or affricate, the point of maximum abduction was not clear. It was not possible to determine the boundary between the devoiced vowel and the preceding voiceless consonant, because the voiceless fricative and voiceless vowel blended into each other so well.

Simada et al. (1991) carried out an electromyographic (EMG) study of the cricothyroid (CT) and sternohyoid (SH) muscles using bipolar hooked-wire electrodes during the production of the devoiced /u/ in Japanese. CT is a broad, paired muscle superficially placed upon the larynx and divided into two muscular groups: the pars recta and the pars obliqua (Hardcastle, 1976: 80). It applies longitudinal tension to vocal folds, and is active during F0 changes (Lieberman and Blumstein, 1988: 120), but lowering of F0 may be a passive movement, simply a cessation of activity in the CT and other muscles

---

<sup>3</sup> A consonant /t/ before /i/ is phonetically realised as an affricate [tɕ].

such as the vocalis and thyroarytenoideus externus which normally increase the tension in the vocal folds (Hardcastle, 1976: 82).

SH is a long, flat, paired muscle, and is usually described as one of the strap muscles of the neck (Hardcastle, 1976: 73). SH lowers the hyoid if muscles that go from hyoid to skull and mandible are slack, and also stabilised hyoid when muscles like digastric tense to open the mandible. It may be active in initiating phonation register shifts (Lieberman and Blumstein, 1988: 120). It also lowers the larynx. Lowering the larynx may serve a number of functions including lowering the fundamental frequency, by increasing the supra-glottal pressure during the articulation of voiced lax stops such as [d], [g] (Hardcastle, 1976: 74). There was increased electrical activity in the CT, which correlated with the stressed word, also there is a peak in the CT activity, which is correlated with the peak in fundamental frequency for the stressed word (Lieberman and Blumstein, 1988: 104).

Simada et al. found different muscular control involving the voiceless [ɥ̥] depending on types of preceding consonants. When a plosive preceded the vowel [ɥ̥], a clear suppression was found in the CT activation, whereas when a fricative or affricate preceded [ɥ̥], the CT did not show any suppression. The SH activity more or less overlapped with that of the CT, but the SH was activated about 20 ms later than the CT and did not cease immediately after the CT activity had stopped. This muscular activity was observed when a voiceless plosive preceded the /u/. However, this was not the case when [ɥ̥] was preceded by a fricative or affricate. Therefore, Simada et al. divided vowel devoicing into two groups depending on the preceding consonants: (a) when a devoiceable vowel occurs after a plosive, it is weakened or becomes voiceless, and (b) when a devoiceable vowel occurs after a fricative or affricate, it is usually deleted or reduced to a mere portion of the preceding consonant.

Both studies by Weitzman et al. and Simada et al. suggested that when a high vowel was preceded by a voiceless fricative, or an affricate, the consonant and the vowel became closely coarticulated and the movements of the laryngeal muscles were integrated with each other. When a high vowel was preceded by a voiceless plosive,



however, the laryngeal muscular activities worked separately for the voiceless plosive and the voiceless vowel. In other words, when a vowel is preceded by a fricative, it is physiologically more likely for the vowel to undergo a devoicing process than when a vowel is preceded by a plosive.

The problem is that vowel devoicing rates in the present study differed significantly when vowels were preceded by fricatives from when preceded by affricates. Preceding affricates showed similar devoicing rates to preceding plosives rather than to fricatives in the condition including utterance final positions and consecutive devoicing environments. This may have been caused by the environments in which vowel devoicing preceded by affricates occurred. In the consecutive devoicing environments, not all devoiceable vowels are devoiced, and some remain voiced, because consecutive devoicing over two morae is not favoured. Since preceding fricatives seem to trigger more devoicing than any other type of consonants, vowels preceded by affricates are likely to remain voiced when vowels in adjacent morae are preceded by fricatives in the consecutive devoicing environments. Also quite a few vowels preceded by affricates appeared in a word final position and positions followed by an internal word boundary, where vowel devoicing is not favoured (see Table 5.11, in Chapter 5, Section 5.4.5) As mentioned earlier in this chapter, a slow tempo triggered a pause at an internal word boundary. These factors may have brought the devoicing rates for preceding affricates down.

Secondly, considering the case of devoicing rates in consecutive devoicing environments when the preceding consonants are plosives or affricates, there was a significant effect of speaking rates. It may be that speaking rate affect the constraints for consecutive vowel devoicing as discussed in Chapter 5: (1) devoicing of more than two consecutive morae is not favoured, and (2) when there is no blocking factor in a word, voiced high vowels in consecutive devoicing sites are not favoured, but they are allowed if a blocking factor of devoicing is present. These constraints were based on reading a list of citation forms of words at a normal speaking rate. At the fast tempo, these blocking factors may become less effective and allow more devoiceable vowels to undergo the devoicing process consecutively. At the slow tempo, on the other

hand, the blocking factors may become more effective and block more devoiceable vowels from becoming voiceless. It should also be noted that at the slow tempo, the speakers often paused at an internal word boundary making a preceding vowel word final, which is one of the blocking factors.

However, in single devoicing environments and when devoiceable vowels were preceded by fricatives, the speaking rate had no significant effect. This implies that vowel devoicing in Japanese is not the result of a mere glottal assimilation between a short vowel and adjacent voiceless consonants, nor an optional fast speech phenomenon. If it was caused mainly by glottal assimilation simply skipping phonation for the vowel, there would be no way to explain why the devoicing rates did not differ significantly, depending on the speaking rates, when consecutive devoicing environments were included, and why vowels preceded by fricatives showed a high devoicing rate even at the slow tempo. There are two factors relating to the effect of speaking tempo on vowel devoicing; firstly to types of preceding consonants and secondly to the devoicing condition. If the first devoicing factor is tempo sensitive, it defers to the second factor, but if the first factor is tempo non-sensitive, the vowel becomes voiceless by default, as shown in Figure 6.15.

**Figure 6.15      Two factors relating to the effect of speaking tempo**

<div>Tempo effect</div> <div>Factors</div>	Tempo sensitive		Tempo non-sensitive	
Type of preceding consonants	plosives  (affricates)		fricatives  (affricates)	
Devoicing condition	single	consecutive	single	consecutive
Devoicing status	almost always	very rare	(almost) always	sometimes in certain environments

## 6.5 Conclusion

Vowel devoicing in Japanese does not only occur in fast speech. It is a phonologically controlled process occurring also in normal and slow speech. However, the experimental results showed that vowel devoicing is also partly a phonetic process; although the effect of speaking rate was generally minimal, it was effective in certain devoicing environments. As for types of preceding consonants, fricatives triggered devoicing consistently with no significant effect of tempo found for preceding fricatives in any conditions. For preceding plosives and affricates, speaking tempo became effective in consecutive devoicing environments, but not in single devoicing environments, where most of the devoiceable vowels became voiceless.

In consecutive devoicing environments when preceding consonants were either plosives or affricates, the effect of speaking rate was significant. The vowels were devoiced more at faster tempi than slower tempi, which created many examples of consecutively devoiced vowels over two morae. As for non-high vowel devoicing and high vowel devoicing in non-devoicing environments, they occurred more at faster tempi than slower tempi. High vowel devoicing as a whole does not seem to be a fast speech phenomenon, but consecutive devoicing over two morae (non-high vowel devoicing and high vowel devoicing) in non-devoicing environments seem to be fast speech phenomena.

However, detailed analysis of devoicing rates suggested that speaking rate may not have been a direct controlling factor for vowel devoicing, but other factors which devoicing is directly affected by, such as position in an utterance and presence of an internal word boundary, are sensitive to speaking tempo and these factors may directly trigger voicing of vowels in the devoicing environment.

It should also be noted that the low devoicing rates with preceding affricates may have been caused by the phonological environments of the test words in the experiment, and devoicing rates could be as high as when vowels are preceded by fricatives, as physiological studies suggest.



The experimental results again showed that Japanese vowel devoicing is conditioned by both phonetic and phonological factors. Fricatives preceding a vowel create a phonetically ideal environment for devoicing, therefore the vowel is almost always devoiced even at a slow tempo. Devoicing, though, became tempo sensitive when a vowel was preceded by plosives. The effect of speaking tempo was effective in the consecutive devoicing environment. When a tempo increased, the incidence of consecutive devoicing over two morae increased, whereas at the slow tempo, only certain devoiceable vowels became voiceless. The results implied that although speaking tempo can have a positive effect on devoicing, devoicing is still strongly conditioned by a phonological factor, namely single and consecutive devoicing environments.

Vowel devoicing may be part of the overall process of vowel weakening, and its final state is completely voiceless. The amplitude of vowels in devoicing environments will be compared with vowels in non-devoicing environments in Chapter 8.

## Chapter 7

# Duration of morae with a devoiced vowel

### 7.1 Introduction

Standard Japanese is often cited as a 'mora-timed language'. However, 'mora' as well as 'syllable' are not very well defined. As discussed in Chapter 3, Section 3.2, the theory that 'mora' is the basic unit of Japanese rather than syllable is disputable, though there are some theoretically convincing evidence to support the 'mora', such as the theory of accentuation, examples of speech errors and the results of a word-blending experiment and a language game called 'shiritori (hip-taking)'. The patterns of accentuation show that lexical accent location is based on the moraic unit, not the syllabic unit. In an experiment using a language game 'shiritori', Katada (1990) found that even preliterate Japanese speaking children used the mora as a unit of Japanese, rather than the syllable. Another possible reason for supporting 'mora' as a basic unit of Japanese is the duration of the morae. It has been traditionally thought that each mora in Japanese has more or less the same duration. Numerous studies have been carried out to examine durational characteristics of the morae with respect to production and perception, but their results have never agreed unanimously.

Some studies (H. Sato, 1978, Higuchi et al., 1978) have suggested that mora-timing is a perceptual phenomenon rather than a feature of production of Japanese speech,

but the majority of works on mora-timing have been based on speech production. The present study will also be in terms of production.

Segmental duration is affected by various factors. Sagisaka and Tohkura (1984) and Takeda et al. (1989) found crucial controlling factors for Japanese segmental durations; these were (a) type of phonemes, (b) neighbouring phonemes, (c) mora position in a breath group and (d) speaking rate. The dominant factor for word-level durational control was the effect of neighbouring phonemes, and the second most crucial factor was the mora count of the word uttered in isolation. Utterance-final lengthening and utterance-initial shortening were also reported as factors.

There are quite a few studies which support the mora as a basic unit of Japanese in terms of the durational hypothesis. Among those supporting the mora as a basic unit from a durational point of view, Han (1962a) carried out a spectrographic study comparing the duration of syllables with one or two morae, and found that the duration of the syllables was proportionally related to the number of morae in the syllables. She measured the duration of words with the same number of syllables but differing number of morae, such as /i/ 'stomach' vs. /i.i/ ([i:]) 'good'<sup>1</sup> and found the durational ratio of short vowel to the long vowel to be approximately 1 : 2. When the short vowel was preceded by a voiced consonant as in /bo/ and /bo.o/, the durational ratios between a short vowel and a long vowel was 1.0 : 2.5, and when the preceding consonant was voiceless as in /se/ and /se.e/, it was 1 : 3. However, when the durations of the entire syllables were compared, the duration of /bo.o/ and /se.e/ was approximately twice as long as the duration of /bo/ or /se/ respectively.

Han also measured the closure duration of single consonants and their double counterparts, e.g. /t/ in /i.ta/ 'was present' vs. /tt/ in /i.t.ta/ 'went', /k/ in /ha.ka/ 'grave' vs. /kk/ in /ha.k.ka/ 'peppermint', etc.<sup>2</sup> Her results showed that the durational ratio

---

<sup>1</sup> A mora boundary is marked with a symbol /. /.

<sup>2</sup> Japanese has voiceless double consonants, such as *sakka* /sa.k.ka/ 'writer', *icchi* /i.t.ti/ 'accordance', *issitsu* /i.s.si.tu/ 'one room', etc. They are phonetically realised as long consonants. Voiced double consonants do not occur in conservative accents, but they occur in recent loan words, e.g. *beddo* /be.d.do/ 'bed', *beekon eggu* /be.e.ko.N.e.g.gu/ 'bacon and egg'. (see Chapter 2)



between the closure of single consonants and that of their double counterparts is from 1: 2.6 to 1: 3.0. Similar to the cases of the long vowels, the entire durations of /t.ta/ and /k.ka/ were twice as long as those of /ta/ and /ka/ respectively. Han concluded that since both syllables with long vowels and syllables with double consonants were more than twice the length of their short counterparts, the duration of a segment was not the basic unit of duration : the unit of duration in Japanese was associated with the mora and the actual duration of each mora was approximately the same. She also noted that there was, within a moraic unit, a strong tendency for the components of the moraic unit to balance each other in order to obtain equal duration with neighbouring units.

Hoequist Jr. (1983) also studied the duration of morae in various environments, and found similar results to Han's, although he obtained slightly different mora duration ratios. Hoequist Jr.'s data from real words showed the ratio between /CV/ and /CV.V/ as 1 : 1.7, and the data from nonsense words showed a ratio of approximately 1 : 2, which was similar to Han's results. He claimed that the unit of duration corresponding to the mora was less than the entire syllable, but more than just the acoustic realisation of the vowel. He concluded that Japanese required comparatively strict durational control, and that duration in Japanese was tied closely to the structural features of morae and seemed to keep mora durations from varying much.

A less strict version yet still supporting durational mora-timing is a study by Port et al. (1980). Port et al. compared the durational compensation processes of Japanese and Arabic, and found that although not perfectly, durations of Japanese vowels varied inversely with the duration of an adjacent consonant, and the consonant constriction duration varied inversely with the inherent vowel durations, whereas very little evidence of timing compensation was found in Arabic. However, Port et al. cast doubts on temporal compensation within a mora. Instead, they suggested that temporal compensation extended across several morae so that the word consisting of the same number of morae have a nearly constant duration inspite of varying durations of individual segments. They also suggested that a basic durational unit in

Japanese was not a /CV/ unit, but might be something like the spacing of vowel centres.

Y. Sato (1993) also studied the durations of syllable-final nasal and syllable-initial nasals in Japanese, Korean and English. In Japanese, the syllable-initial nasal and the syllable-final nasal have phonologically different functions: a syllable-initial nasal is always followed by a vowel and together with the vowel constitutes a mora, but a syllable-final nasal occurs at the coda position of the syllable, and it constitutes a mora on its own. Her results revealed that in Japanese a syllable final nasal, which is moraic, is significantly longer than the syllable-initial nasal, which is non-moraic, but such durational tendencies were not observed in Korean or English in the same environment as in Japanese. Sato concluded that there was a strong tendency for Japanese to try to equalise the duration of morae.

Another less strict version comes from Campbell and Sagisaka (1991) and Campbell (1992). They studied the segmental durations of Japanese read sentences. Both studies failed to find equal timing of morae in raw durations, either, but found mora-based segmental elasticity when they compared the z-scored durations of segments, supporting the mora-timing hypothesis. Since the inherent duration of segments differs, they used the normalised measure of segmental durations, i.e. z-scores. Z-score is a statistic which shows how far the obtained figures are from the mean distribution; i.e. it can be used as a measure of distance between the obtained figures and the mean (Hatch and Farhady, 1982: 82-89).

The results of Campbell and Sagisaka and Campbell showed that (a) preceding consonants, (b) following consonants and (c) a vowel in a following mora were the main factors which have a significant effect on duration at the segment level. Another observation was that in Japanese speech it is the consonants that have the greatest effects on speech timing, and that the vowels show less variation both in quality and in durational characteristics. It was also suggested that durational compensation worked within CV sequences (a moraic unit) rather than V-C sequences (i.e. across a mora boundary): the results showed (1) negative correlation

between the durations of C and V in a CV sequence but positive correlation between the durations of C and V in a V-C sequence, and (2) in a CV unit, the vowel was much shorter after an unvoiced stop than after a voiced stop, voiceless stops being inherently longer than voiced stops. Long vowels were not twice the duration of their short counterparts, but about 50% longer. An interesting finding was that consonants preceding long vowels in the same syllable increased their duration approximately 50% as well. That suggests that the preceding consonants increased their duration in order to make up for the insufficient durational compensation of the following long vowels. Also from the evidence of the moraic nasal /N/, double consonants and /CjV/ morae, which clearly took up a mora-sized slot in the timing framework, their study concluded in support of the hypothesis that the mora, and not the syllable, was the prime unit of timing in Japanese.

However, not all studies agree with this view. Beckman (1982) showed that morae could have differing durations and questioned the phonetic reality of a mora as a unit of equal duration. Beckman studied the duration of morae of varying phonetic composition: (1) voiceless obstruents with a following high vowel which is likely to be devoiced or deleted<sup>3</sup>, (2) contrasting short and long consonants, (3) vowels in various consonantal environments, (4) consonants in various vocalic environments, and (5) in words differing only in accent pattern. None of her results support equal duration of morae. There was some tendency towards temporal compensation within CV morae, but it might be attributable to universal phonetic phenomena rather than a language specific feature such as mora-timing. Beckman suggested that the reality of morae is traceable to a mora-based orthography.

Other studies presented results challenging the phonetic reality of morae. Higuchi and Fujisaki (1980) and Sagisaka and Tohkura (1981) (cited by Kawasaki, 1983: 394) studied segmental durations and denied the mora-based durational compensation. Higuchi and Fujisaki examined segmental durations in the sequence

---

<sup>3</sup> At the time of her papers (1982 and 1984), Beckman defined 'vowel devoicing' as psychologically 'devoiced' but phonetically 'deleted'. This definition was later revised (see Chapter 7).



/C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>2</sub>C<sub>3</sub>V<sub>3</sub>/ of varying composition of consonants and vowels, and found that the durational compensation between segments worked between C<sub>1</sub> and V<sub>1</sub>, C<sub>2</sub> and V<sub>2</sub>, or C<sub>3</sub> and V<sub>3</sub> (within a mora), but the durational compensation between units, i.e. /V<sub>1</sub>C<sub>2</sub>/ - /V<sub>2</sub>C<sub>3</sub>/ (a /V-C/ unit, across a mora boundary) was greater than that between /C<sub>2</sub>V<sub>2</sub>/ - /C<sub>3</sub>V<sub>3</sub>/ (a /CV/ unit, within a mora): the longer the /V<sub>1</sub>C<sub>2</sub>/, the shorter the /V<sub>2</sub>C<sub>3</sub>/, and vice versa. The study on durational relationship between segments by Sagisaka and Tohkura also found similar results: greater compensation between the /V-C/ sequences than the /CV/ sequences.

Beckman's study as well as other earlier studies on the mora-timed hypothesis was later criticised by Port et al. (1987) on the ground that Beckman's idea of a mora simply applied to the conventional view of the mora as taught in Japanese schools. Port et al. did not reject the measurements of segments, morae or syllables obtained by other studies, including Beckman's, but studied the durational features of the morae in terms of larger units than segments or morae. They demonstrated that the duration of each mora was not necessarily equal, but the number of morae in a word determined the duration of words. Words with an increasing number of morae increase in duration by nearly constant increments. In other words, two-mora words are roughly twice as long as one-mora words, and three-mora words are roughly three times as long as one-mora words, and so on. They claimed that segmental duration was not the most important feature of the mora-timing hypothesis, but had to be discussed in relation to the overall duration of a word.

Port et al. (1987) conducted detailed phonetic studies to examine the reality of mora-timing. They investigated the relationship between (i) word duration and the number of morae in the word, (ii) speaking tempo and the word duration, (iii) segmental features and mora-timing, and (iv) segmental and word durations and the number of morae and syllables. Their experimental results showed that the duration of each mora was not necessarily equal, but the number of morae in a word determined word duration, regardless of segmental content, presence of devoiced vowels, or the number of syllables in the word. Words with an increasing number of morae increase in duration by nearly constant increments. In other words, two-mora words

are roughly twice as long as one-mora words, and three-mora words are roughly three times as long as one-mora words, and so on. They claimed that segmental duration was not the most important feature of the mora-timing hypothesis, but had to be discussed in relation to the overall duration of a word.

In their first experiment, they measured the duration of whole words. They used two sets of nonsense words differing in the number of morae from 1 to 5 as listed in (1a) and (1b), in which and hereafter mora boundaries are marked with the symbol “.”.

(1)	<u>1-mora</u>	<u>2-morae</u>	<u>3-morae</u>	<u>4-morae</u>	<u>5-morae</u>
a.	/ra/	/ra.ku/	/ra.ku.da/	/ra.ku.da.ga/	/ra.ku.da.ga.si/
b.	/si/	/s <sub>i</sub> .ta/	/s <sub>i</sub> .ta.ku/	/s <sub>i</sub> .ta.k <sub>u</sub> .su/	/s <sub>i</sub> .ta.k <sub>u</sub> .su.ru/

The test words were embedded in a carrier sentence “*Kore wa \_\_\_\_\_ desu*”. The underlined high vowels in the second set of test words listed in (1b) can be devoiced because they are surrounded by voiceless consonants. Four native speakers of Japanese pronounced the test words in the carrier sentence at a comfortable tempo.

Their results showed that the duration of whole words pooled across speakers for both sets showed similar durations dependent on the number of morae in the word, despite the different segmental content in each set, and the devoicing of some vowels in the second set of words. Word duration increased by roughly 120 ms, which was the average duration of a mora. In other words, each additional mora increased the total word duration by nearly an equal amount.

In their second experiment, Port et al. tested the mora-timing hypothesis at different speaking tempi, using test words which were existing words, differing accentuation, and the number of morae and syllables, but having more or less the same segmental content in each set: e.g. /ta/ ‘field’ - /ta.-ka/ ‘hawk’ - /ta’.-ka.-sa/ ‘height’ - /ta.-ka.-sa.-wa/ ‘high swamp’ - /ta.-ka.-sa.-sa.-ki/ ‘high tree’. The test words with the same number of morae did not always have the same number of syllables. For example, /na.-mi.-da.-gu.-mu/ ‘be moved to tears’ and /hi.k.-ka.-ke’.-ru/ ‘to hang’ are both 5-

mora words, but the former is 5 syllables, whereas the latter is 4 syllables. Similarly, /ka.-ku.-si.-do'.-ko.-ro/ 'hiding place' is 6 morae - 6 syllables, but /hi.k-ka.ke.na'.i/ 'not to hang' is 6 morae - 5 syllables. Five subjects pronounced the test words in a carrier sentence four times each (although only two samples were used for analysis) at slow and fast tempi.

The results showed that the duration of the words pooled across speakers again increased constantly depending on the number of morae, at both tempi, but was not dependent on the number of syllables. There was very little difference between durations of words of the same number of morae, despite large differences in the segmental content of the words. At the slow tempo, the size of increment was larger and there was less deviation in duration for the longer words than at the fast tempo.

The subsequent two experiments investigated (a) the effect of segmental features on mora timing and (b) the durational compensation between segments in relation to the mora and syllable timing. Their results showed that the overall durations of the entire words were very similar despite phonetically different segmental content, and that the evidence of durational compensation between segments was not syllable based but rather mora based.

Their third experiment examined the effect of segmental phonetic features on mora timing. They chose 8 nonsense test words of 2 morae each contrasting vowels /a/ and /u/ in the first mora, and voiced and voiceless consonants in the second mora: /kuka/, /kuga/, /kaka/, /kaga/, /kata/, /kada/, /kasa/ and /kaza/. The /u/ in /kuka/ was usually voiceless because it was between voiceless consonants, and /g/ could variably become velar nasal [ŋ] (see Chapter 3). 10 subjects pronounced the words in a carrier sentence. It is universally agreed that voiceless consonants are intrinsically longer than their voiced counterparts, and high vowels are also inherently shorter than non-high vowels. However, the results showed that the overall durations of the entire words pooled across speakers were very similar, despite phonetically different segmental content. The difference in word duration between the longest /kasa/ and the shortest /kuka/ was only about 7 %.



Their fourth experiment examined segmental duration in relation to the mora and syllable. Four pairs of words were used in the experiment. One pair was 2 morae and 2 syllables (/bu.-ku/ and /ba.-ku/), one pair was 3 morae and 3 syllables (/bu.-ku.-do/ and /ba.-ku.-do/), and the other pairs had 3 morae but 2 syllables: one of them was /CV.V-CV/ (/bu.u.-ku/ and /ba.a.-ku/) and the other was /CV.C.-CV/ (/bu.k.-ku/ and /ba.k.-ku/). The results showed that the words with the same number of morae were similar duration. They found no syllable based compensatory shortening in the duration of the initial /b/: if a syllable was a primary unit for segmental timing, the durations of the initial consonant /b/s in the heavy syllables /CVV/ and /CVC/ would have been shorter than those of the light syllables /CV/. On the other hand, the /k/s following the long vowels in /baaku/ and /buuku/ were longer than the /k/s following short vowels. It appeared that although heavy syllables had an extra mora, they were generally shorter than other 2 morae composed of /CVCV/. Therefore, the longer duration of the /k/ made up for the shorter duration of the preceding morae.

Overall the experimental results demonstrated that the mora is an important unit of timing in Japanese speech, but proposed a different view of mora-timing from the traditional definition. Each mora did not necessarily have equal duration, but the entire duration of a word was decided by the number of morae in the word. In other words, although the mora is a basic timing unit in Japanese, the durational adjustment does not work within a mora, but works at a higher level such as within a word.

In a more recent study, Han (1994) examined (1) the elasticity of segment durations providing temporal compensation, (2) the similarity in word durations among words with the same number of morae, but with different segmental compositions, and (3) the predictability of the durations of whole words from the number of morae in words. Han used 10 pairs of test words, in which each pair of words had the same number of syllables but differing numbers of morae, and consisted of different combinations of consonants and vowels, e.g./su.-pa.-i/ 'spy' - /su.p.-pa.-i/ 'sour', /ki.-

te/ 'coming' - /ki.t.-te/ 'postal stamp', /si.-ke/ 'storm' - /si.k.-ke/ 'humidity', etc. (mora boundaries are marked by “.” and syllable boundaries are marked by “-”).

Han's results (1994) did not find exactly equal durations for each mora, but showed segmental elasticity within and between morae, so that the duration of the whole word attained targeted durations based on the number of morae in the word. The duration of a given segment was longer before a shorter following segment and shorter before a longer following segment. For example, a geminate consonant is longer than a single consonant in Japanese, but the first part of a geminate consonant, which consists of a mora on its own, has a shorter duration than /CV/ morae. Her results showed that the same vowel was 11 % longer when it was followed by a geminate stop than when followed by a single stop, in word internal position and 9% longer in word final position, which was probably because it was necessary for the duration of the vowel preceding a geminate stop to be adjusted in order to compensate for the shorter duration of a geminate consonant, to achieve the target duration of the entire word.

Han's other measurements found that the mean ratio of 2-mora to 3-mora words was 2.00 to 2.99, and that of 3-mora and 4 mora words was 3.00 to 4.01, despite differing intrinsic durations of each segment. On the other hand, syllable durations varied by as much as 50 %. Han's results confirmed the moraic theory proposed by Port et al. (1987) that mora-based durational adjustment works at two levels, (a) at the mora level and (b) at the word level.

Some studies support the mora-timing hypothesis, but there has been little work on the duration of morae in which vowel devoicing occurs. As discussed in Chapter 5, when a high vowel following a plosive is devoiced its phonetic realisation is a long aspiration, or the plosive is released into a homorganic fricative; and when a vowel following a fricative or affricate is devoiced, the vowel is phonetically realised as a continuation of the preceding fricative. When the vowel is devoiced, what happens to the duration of the remaining consonant ('devoiced mora')? Is the duration of a devoiced mora similar to the duration of the same mora with a fully voiced vowel

(i.e. the vowel is devoiced), or similar to the duration of the remaining consonant only without the vowel (i.e. the vowel is deleted)? This question is related to the traditional dispute over whether the high vowel is devoiced or deleted. Vowel deletion instead of devoicing, however, is based solely on durational features of devoiced morae, not taking account of other aspects such as prosodic features or coarticulation with neighbouring segments.

Beckman (1982) carried out a study comparing the duration of morae with fully voiced vowels and morae with devoiced vowels. Beckman asked her subjects to pronounce test words which included a devoiceable vowel, in carrier sentences three times for each sentence in random order. Her subjects pronounced the test words sometimes with a voiced vowel and other times with a voiceless vowel, which enabled her to compare the duration of the same mora with the voiced vowel and with a voiceless vowel. There were 27 out of 115 instances of a speaker maintaining voicing of a devoiceable vowel in one or two of the three tokens for the same word, which made it possible for her to compare 54 durations (two comparisons for each triplet). Her results showed that in only four out of the 54 pairs (26 plosives and 28 fricatives) was the moraic consonant<sup>4</sup> longer than the prevocalic counterpart by at least the length of the vowel in the CV mora. Among the total number of 54 tokens, only 54% of moraic consonants (46% of moraic plosives and 61% of fricatives and affricates) were longer than non-moraic counterparts, and they were not significantly longer. Although some moraic consonants were longer than their non-moraic counterparts, and given that there were some segmentation problems<sup>5</sup>, Beckman rejected the hypothesis of compensatory lengthening of the preceding consonants as

---

<sup>4</sup> Strictly speaking, in Japanese phonology there are only two types of moraic consonants; (1) moraic nasals /N/ such as /ho.N/ 'book' and /ka.N.-ba.N/ 'sign board', and (2) the first part of geminate consonants such as the first /t/ in /i.t.-ta/ 'to go (past)' and the first /k/ /i.k.-ka/ 'one family' (mora boundaries are marked by "." and syllable boundaries are marked by "-"). The term 'moracic consonant' in the context of vowel devoicing implies the absence of the devoiceable vowel, i.e. the vowel is not devoiced but deleted, which is, strictly speaking, not true, as explained below in this section. However, at the time of her paper (1982), Beckman concluded that vowel devoicing was vowel deletion (see footnote 4 in this chapter) and called the remaining consonant a 'syllabic consonant' ('moracic consonant' by my definition). For convenience in this section, I follow her definition and use the term 'moracic consonant' to mean a consonant whose following vowel is devoiced.

<sup>5</sup> For the details of segmentation problems, see Section 6.4.3



a result of vowel devoicing, and concluded that vowel devoicing was in fact vowel deletion.

Despite Beckman's claim, it may not be appropriate to say that the high vowels are deleted. Her other results (Beckman and Shoji, 1984) showed that Japanese speakers could identify /si/ [ɕi] and /sju/ [ɕɯ] when the vowels were voiceless with a high degree of accuracy. This was because the remaining consonant [ɕ] had the colouring of the devoiced vowels: [ɕ<sup>w</sup>] before underlying /u/ and [ɕ<sup>l</sup>] before underlying /i/. In addition to this, Sugito and Hirose (1988) found that a listener could still identify an accent on a devoiced vowel because of the sharp fall of fundamental frequency on the following mora (see Chapter 4, Section 4.5.2). If there is any influence of the devoiced vowel on neighbouring sounds, the vowel portion might still be said to be perceptually present. In other words, the vowel is underlyingly present, not deleted, i.e. it is devoiced. If there is no evidence of compensatory lengthening of the preceding consonant, but phonetic effects of the vowel are still present, it would be more appropriate to term the phenomenon 'durational reduction' of the mora rather than 'vowel deletion'.

Port et al. (1987) and Han (1994) did not measure the durations of devoiced morae, but measured whole durations of words with devoiced morae. Both of their results showed that the duration of whole words were dependent on the number of morae in a word and the presence of devoiced vowels did not affect the duration of whole words.

All studies on durational features of devoiced morae discuss 'compensatory lengthening' of a remaining moraic consonant. However if a vowel is really 'devoiced', which is a change of glottal state from voiced to voiceless, the duration of the devoiced vowel should be retained. If vowel devoicing involves durational reduction as well as the change of glottal state, the duration of a devoiced mora is reduced and acceptance of the presence of the devoiced vowel is dependent on other factors such as accent placement, coarticulation of preceding consonant, and so on.

Some studies claim durational elasticity of segments based on a moraic unit, and others claim durational elasticity of segments which operates in a larger unit such as a word. On the other hand, Beckman's measurements on devoiced morae showed no durational lengthening of the preceding consonants: i.e. the devoiced morae were reduced in duration as if the vowel portion was deleted. All these studies gave equally convincing results. Most studies' durational statements were based on the idea that vowel are 'deleted' not 'devoiced'.

In this chapter, I will investigate two levels of durational adjustments with particular relation to devoiced morae and words with devoiced morae. Principally, the following points will be examined: (1) the duration of devoiced and non-devoiced morae, (2) the duration of a whole word with and without devoiced vowels, and (3) the effect of the number of devoiced vowels on the duration of a word. The question to be answered is whether vowel devoicing involves durational reduction. One of the difficulties in understanding the vowel devoicing process is the partialness and the optionality of devoicing. Even under ideal conditions for devoicing, a vowel sometimes still fails to become voiceless. In order to explain these features of devoicing, it was proposed that vowel devoicing is part of the vowel weakening process. This new approach may explain the partialness and optionality of devoicing, but it is not yet clear how the vowel weakening process is operated in Japanese. In this chapter, I will examine whether the vowel weakening process in Japanese involves durational reduction similar to the reduction processes in other languages.

## 7.2 Experimental hypotheses

Among the studies of segmental duration, Beckman (1982) argued that since morae with a devoiced vowel were significantly shorter than CV morae ('CV mora' henceforth is used to mean a non-devoiced CV mora), the mora did not have phonetic reality in Japanese. The results of other studies, on the other hand, showed some sort of durational adjustment of segments based on the mora. If we

acknowledge that a mora rather than a syllable is basic to Japanese speech from a durational point of view, a morae with a devoiced vowel presents an interesting case, because if the remaining consonant really constitutes a much shorter mora than a /CV/ mora, it is an exception to mora-based segmental elasticity. Many studies reported mora-based durational adjustment of segments; one would expect it also to operate in the morae with a devoiced vowel, or are they simply durational exceptions? How can Beckman's findings that the mora does not have phonetic reality be accounted for in the studies that found some durational adjustment? If devoiced morae have the duration of only the remaining consonant as examined by Beckman, especially in words with more than one devoiced vowel, vowel devoicing will create a series of shorter morae in a word. In these cases, is it still possible to adjust mora-based duration at the word level and for word duration still to maintain the target duration according to the number of morae in a word?

Some studies suggested no durational reduction of devoiced morae, but others showed reduction. Among studies which showed some reduction, some results suggested durational adjustment at a mora level, but others showed it at a level of larger units such as the word. Perhaps there is adjustment at one level in some situations and not others.

If a vowel is 'devoiced', the duration of the devoiced mora should be more or less equal to a /CV/ mora. However, if vowel devoicing involves durational reduction as well, moraic consonants will be shorter than /CV/ morae. If there is durational adjustment at the mora level as reported by Hoequist Jr. (1983), Y. Sato (1993), Han (1994), etc., the duration of each mora should be more or less equal. If this mora-based durational adjustment also works for devoiced morae, the duration of devoiced morae should be similar to the duration of other morae as shown in Figure 7.1 (a). On the other hand, if mora-timing works at the level of a larger unit than a mora, such as the word as proposed by Port et al. (1987), the duration of devoiced morae does not have to be similar to the durations of other non-devoiced morae in the same word, as the shorter duration of devoiced morae is adjusted for at the word level. Therefore, the durations of devoiced morae could be shorter than other morae like



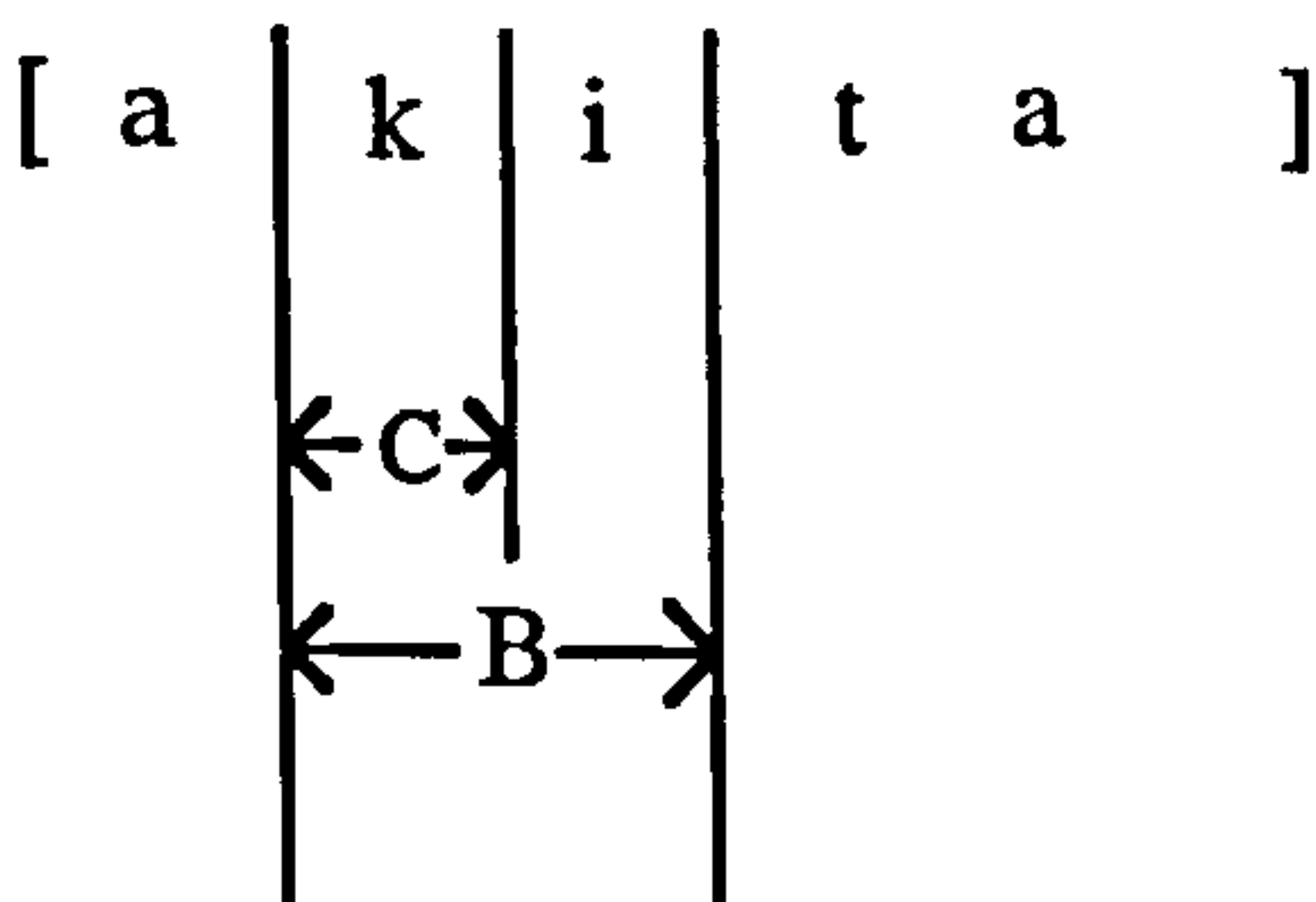
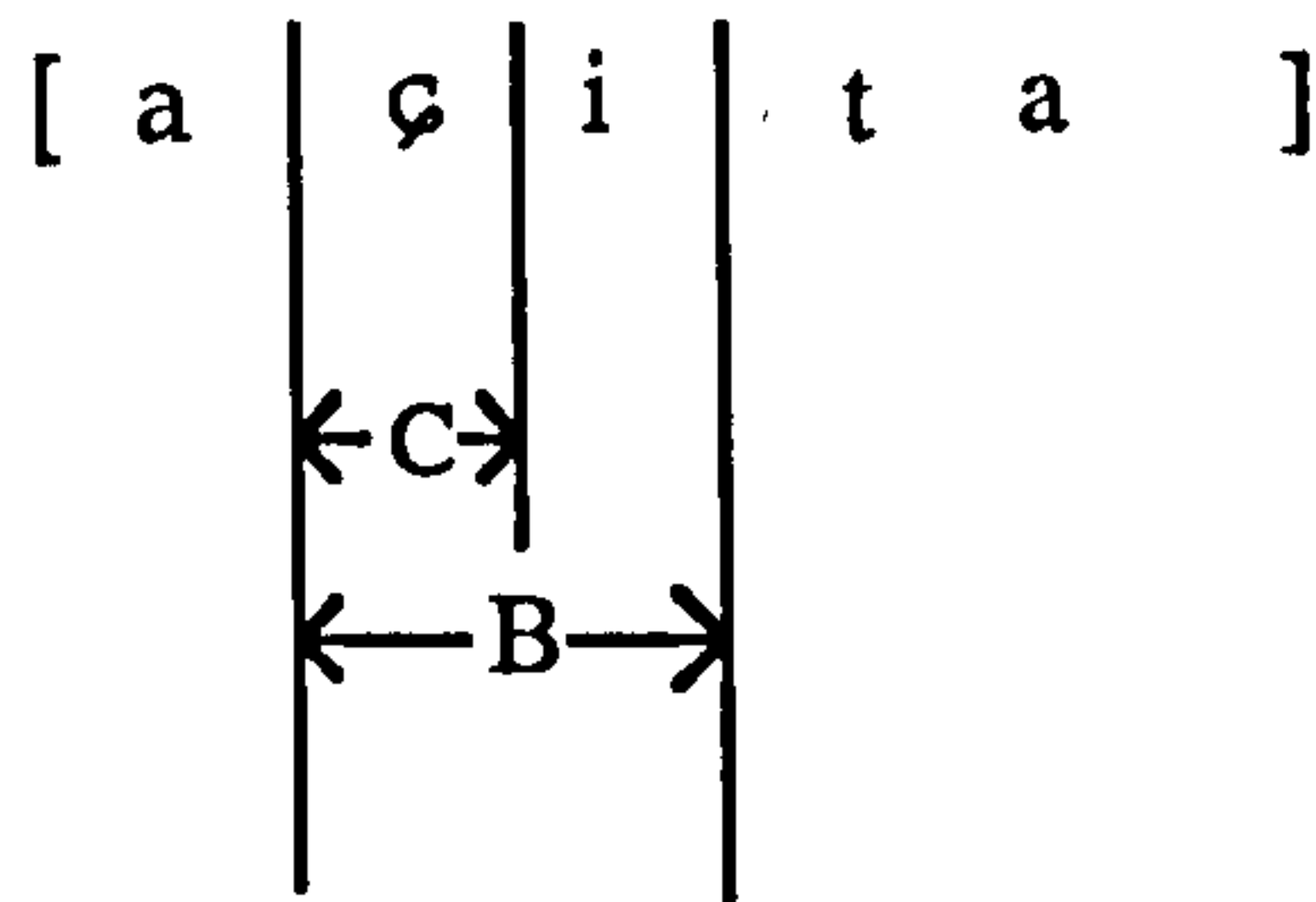
Figure 7.1 (b) as reported by Beckman (1982). Alternatively, the durational adjustment operates at both mora and word levels.

**Figure 7.1** Schematised diagrams of durational analyses of words with a devoiced mora in two devoicing environments: (i) a devoiced vowel is preceded by a fricative, and (ii) a devoiced vowel is preceded by a plosive (mora boundaries are marked by the symbol /<sub>•</sub>/ and voiceable vowels are written in bold italics.)

- (i) preceded by a fricative: *ashita* /a.*si*.ta'/ 'tomorrow'
- (ii) preceded by a plosive: *akita* /a'.*ki*.ta/ 'be bored (past)'

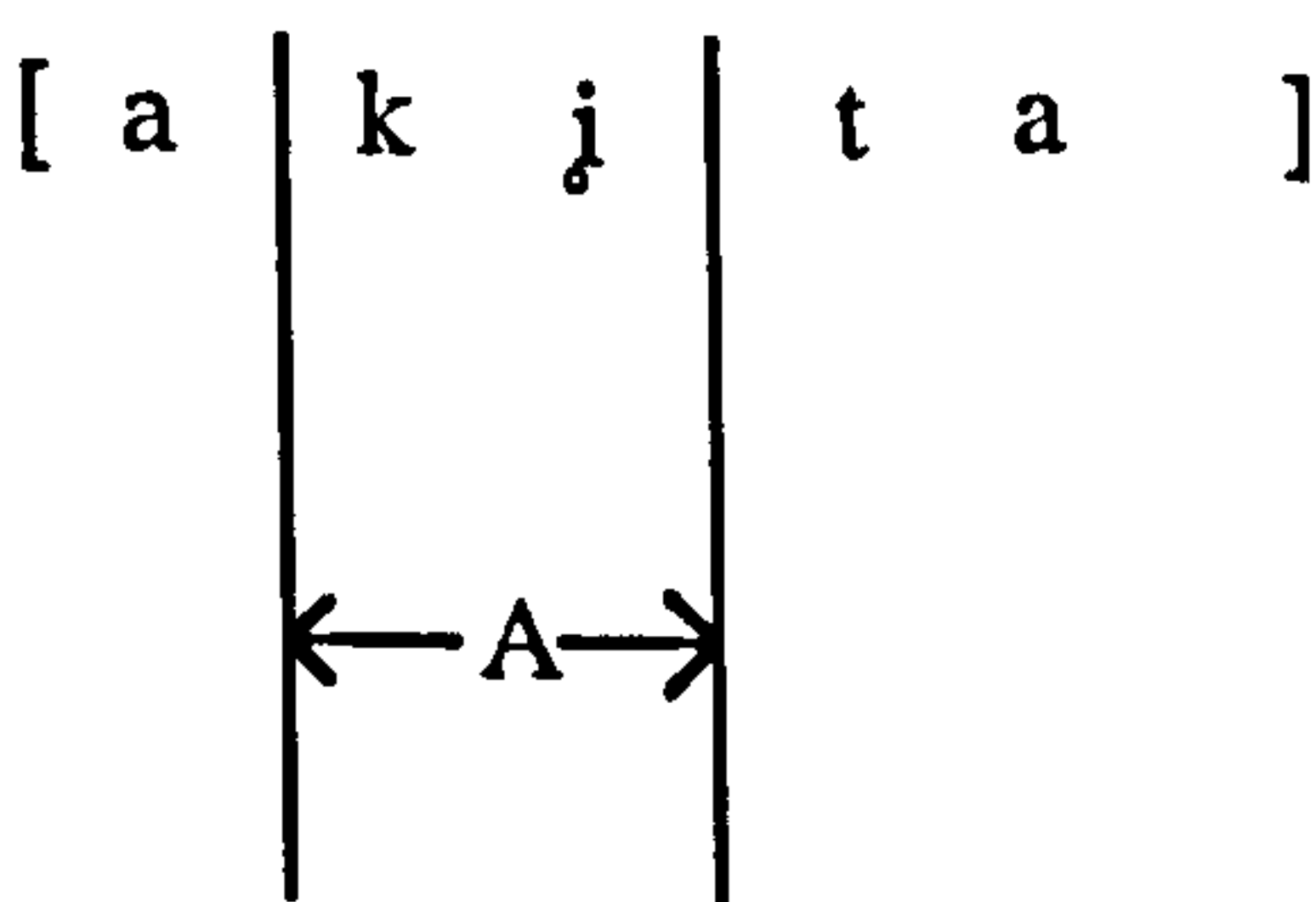
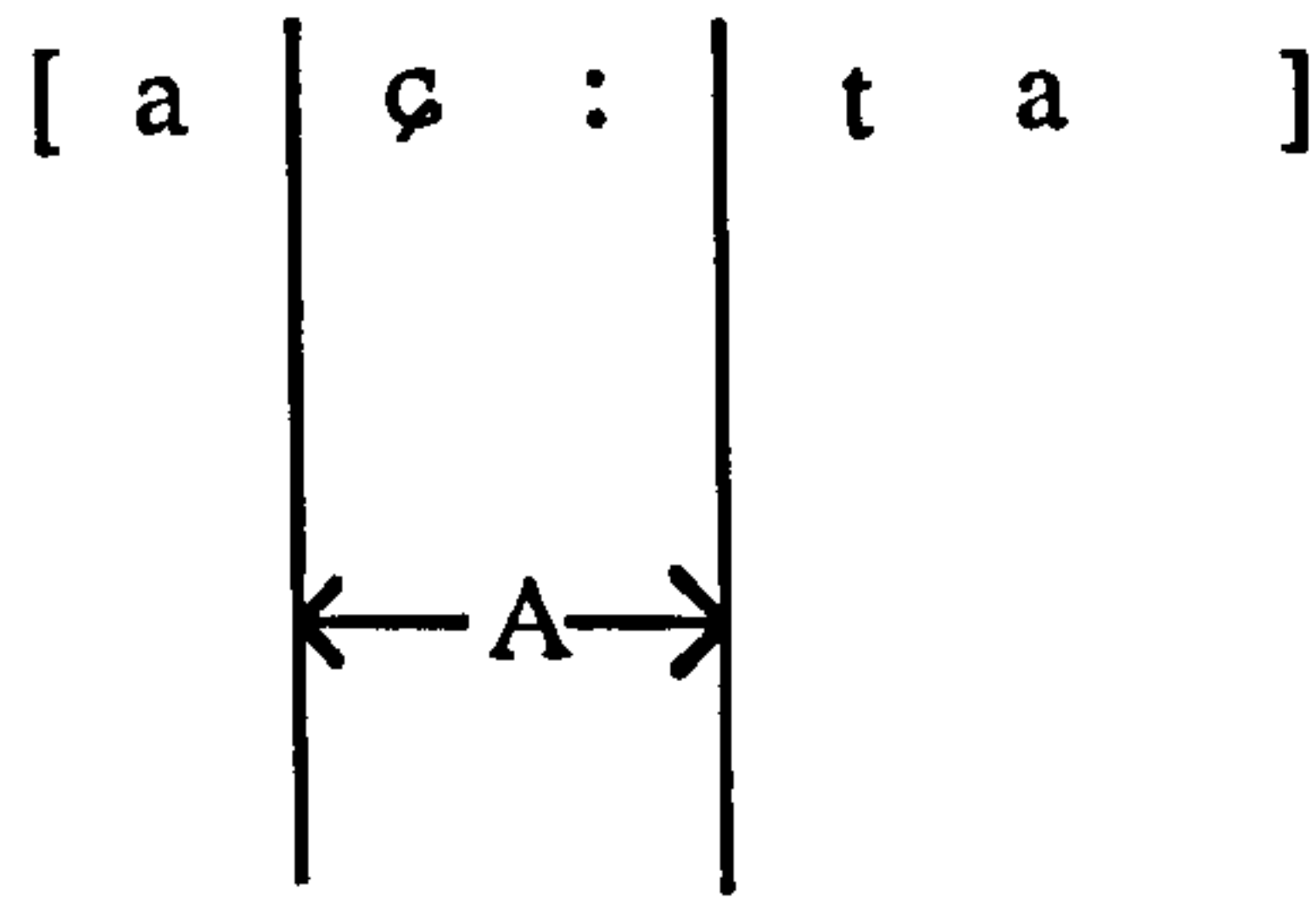
(i) /a.*si*.ta/ 'tomorrow'

(ii) /a.*ki*.ta/ 'be bored (past)'



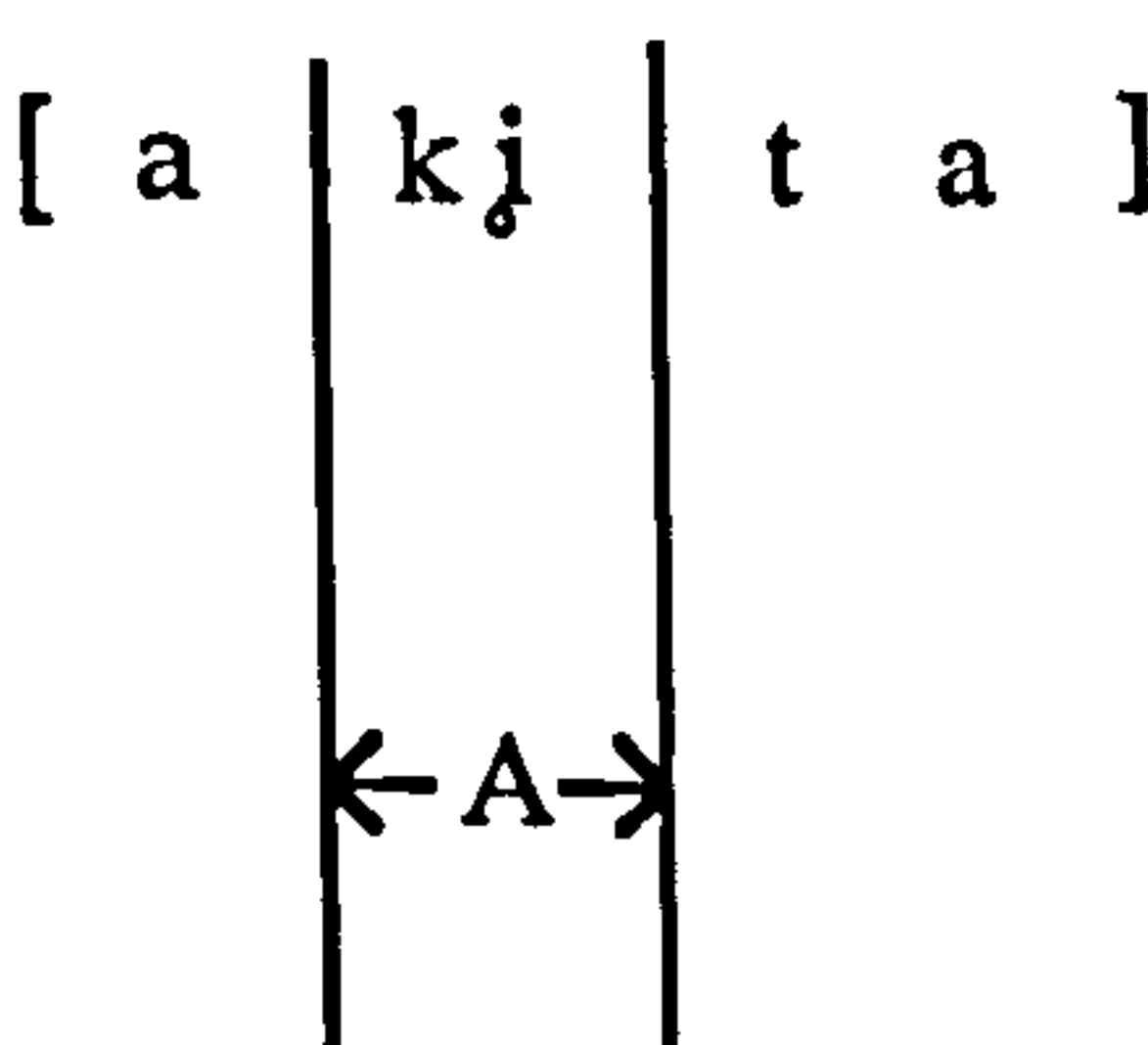
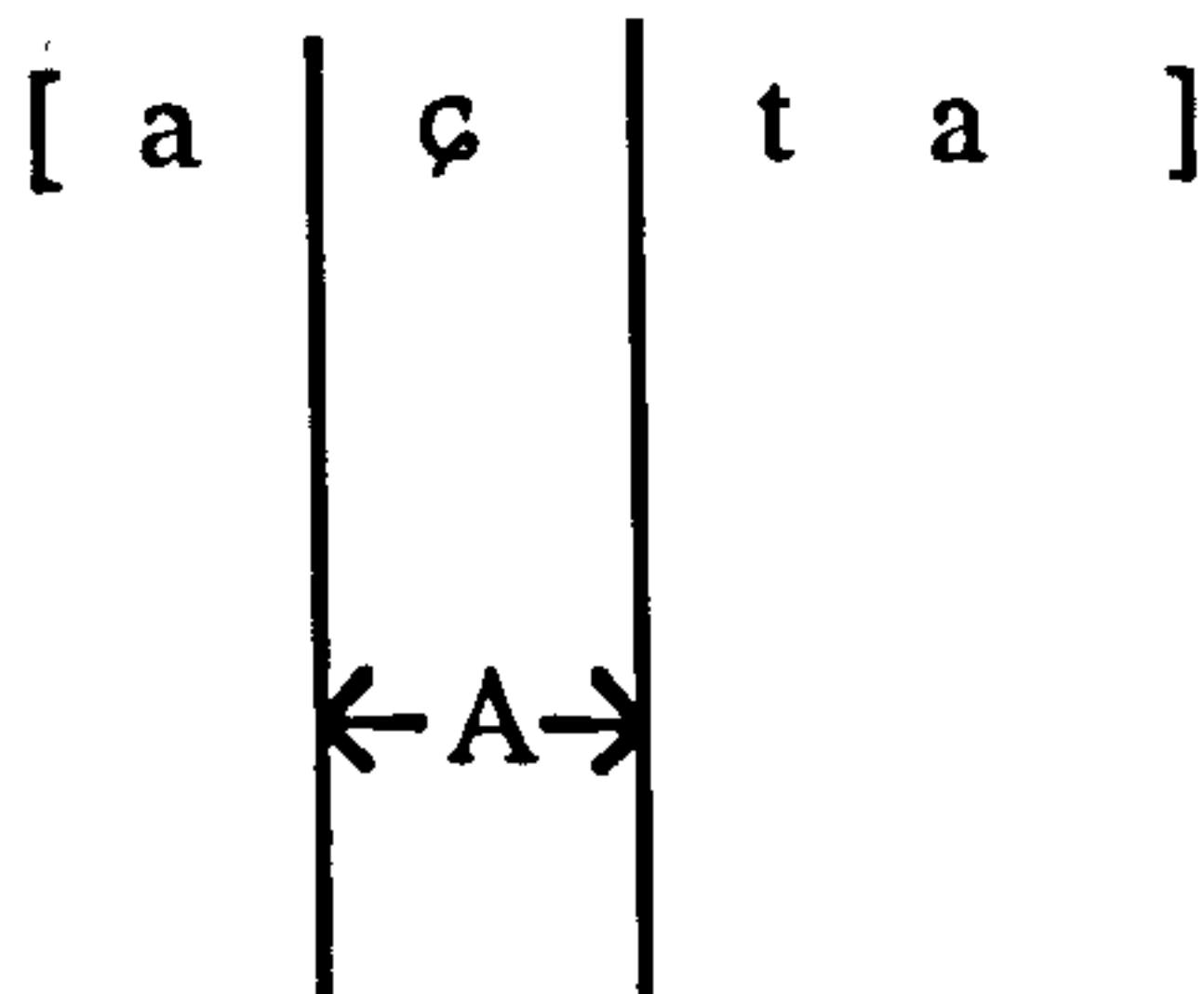
(a) mora-based durational adjustment hypothesis

(A = B)  
(A > C)



(b) word-level durational adjustment hypothesis

(A < B)  
(A > C)



Physiological investigation of vowel devoicing in Japanese showed that devoicing following plosives on the one hand, and following fricatives on the other, involved slightly different laryngeal muscular movements (Weitzman et al., 1976, and Simada et al., 1991) (see Chapter 6, Section 6.4.7). When a preceding consonant is a fricative or affricate, the laryngeal muscular activities of a preceding consonant and a devoiced vowel blend into each other, but when a preceding consonant is a plosive, those muscular activities of plosives and devoiced vowels are still separable. Therefore, the durations of devoiced vowels after fricatives might be shorter than the durations of devoiced vowels after plosives. The durations of devoiced vowels after affricates might come to the intermediate value between moraic fricatives and plosives.

If vowel devoicing involves durational reduction of a devoiced mora, the durations of other segments must be adjusted within a mora as reported by Han (1994), so that the duration of a whole word remains the same as a word of the same number of morae with no devoiced mora. When there are several devoiced morae in a word, and if there is no durational adjustment in each devoiced mora, this creates successions of shorter morae, and the mora-based durational adjustment at the word level does not work, as the series of shorter durations of morae are not recoverable at the word level. Therefore, it is plausible that the greater the number of devoiced morae in a word, the higher the degree of durational adjustment.

Guided by the results of other studies on segmental duration and the mora-timing hypothesis, I will hypothesise the following three for the experiment:

- (i) When a high vowel is devoiced, its process involves not only glottal change but also durational reduction. Therefore, the durations of devoiced morae are shorter than their /CV/ counterparts.

- (ii) If vowel devoicing involves durational reduction, there is a difference in the degree of durational adjustment of the preceding consonants, depending on their types. When a vowel following a plosive is devoiced the duration of the remaining consonant is closer to Figure 7.1 (a) above, whereas when a vowel follows a fricative, the durational ratio is rather like Figure 7.1 (b), and the duration of a preceding affricate is between those of the plosive and the fricative.
- (iii) If vowel devoicing involves durational reduction, mora-based durational adjustment operates both within a mora and at the word level. Therefore, the degree of durational adjustment within a mora is higher when there are more devoiced vowels in a word.

### 7.3 Experimental methods

#### 7.3.1 Subjects and materials

The same six speakers for the experiments in Chapters 5 and 6 were the subjects for this experiment. The data analysed for Chapter 5 were again used for the experiment.

In order to minimise the effects of various factors which control segmental duration such as (a) type of phonemes, (b) neighbouring phonemes, (c) mora position in a breath group, and (d) speaking rate (see Section 7.1), durational comparisons were made only in the same mora in the same word uttered by the same speaker in utterance internal positions. Word final morae were excluded from the comparison, even if there was variation in voicing of the vowel. Some test words had possibilities of consecutive devoicing and devoicing sites occurring word initially, but there were no cases of voicing variations in word-initial position. There were some words in which the subjects inserted a pause after an internal word boundary, and the preceding mora had voicing variation of the vowel. In those cases, the words were excluded from the data as the preceding mora became utterance-final allowing for the possibility of utterance-final lengthening, and the following mora became word-initial in which case there could have been utterance-initial shortening.



### 7.3.2 Procedure

In order to compare the duration of morae with a voiced vowel and with a devoiced vowel, I used a method similar to that of Beckman (1982) (see Section 7.1). In my experiment, when the six subjects pronounced the 41 test words with 74 devoicing sites three times each in random order, their pronunciation of devoiceable vowels in the same words was not always consistent: if a vowel was not either consistently devoiced or consistently voiced, in some cases the vowel was devoiced once and voiced twice, in other cases the vowel was devoiced twice and voiced once (see Chapter 5). When there was variation in voicing of the same devoiceable vowel in the same word, the word was segmented and the segmental durations were measured. The duration of moraic consonants was compared with that of corresponding CV morae. The comparison was only made between utterances by the same subject.

In Chapter 5, examples of voiced and devoiced vowels in devoicing sites were presented (Chapter 5, Figures 5.3a, 5.3b, 5.4a, and 5.4b). However, vowels in devoicing sites were not simply divided into two glottal states, voiced and voiceless. There were quite a few marginal cases, i.e. partially voiced vowels (Chapter 5, Figures 5.5a and 5.5b). The phonetic reality of vowel devoicing is not a clear-cut distinction between voiced and devoiced vowels, but a gradual change of glottal state from fully voiced to completely voiceless. Vowels in devoicing sites can be at any state between fully voiced and completely voiceless (see Chapter 5, Figure 5.8). However, ‘partially voiced’ vowels were categorised as ‘voiced’ vowels if there were three or more periodic waves shown on the waveform, as it is very difficult to draw an absolute boundary between fully voiced and partially voiced vowels.

The data were collected from the same tokens used in Chapter 5: 738 recorded words (41 test words x 3 times x 6 subjects) containing 1332 devoiceable vowels (74 devoicing sites x 3 times x 6 subjects). Among them, 45 devoicing sites (45 sites x 3 times = 135 high vowels) had voicing variations, excluding word-final position and pre-pausal position. One female subject MT did not show any voicing variation. She either constantly voiced or devoiced the same vowel in the same devoicing site.

Therefore the results do not include her data. All words which had voicing variation were segmented using Waves+ speech analysis on a SUN workstation.

### 7.3.3 Durational measurements and segmentation criteria

Two sets of measurements were taken for each devoicing site: for example, if a vowel in a word was devoiced in one utterance (p) and voiced in the other two utterances (q) and (r), two ratios (p/q) and (p/r) were calculated; if a vowel was devoiced in two utterances (x) and (y) and voiced in the other utterance (z), two ratios (x/z) and (y/z) were obtained.

Segmentation was primarily based on waveforms, and secondarily on spectrograms. Segmentation criteria were as follows:

- (1) The duration of vowels was determined from the point where the periodic waveform started until the point at which the periodic waveform ended.
- (2) The beginning of the periodic waveform usually corresponded to the beginning of the first formant on the spectrogram, but clear striations of vowels on the spectrogram very often ceased before the periodic waveform ceased on the waveform. In these cases, the segmentation was based on the end of periodicity in the waveform (see Section 7.4.3 for further discussion).
- (3) The duration of fricatives was taken from the beginning of the aperiodic waveform to its end. There were no cases of fricative sequences in the test words.
- (4) Separate measurements for closure and aspiration were taken for plosives, and similarly separate measurements for closure and frication periods were taken for affricates, except voiceless plosives and voiceless affricates in word initial position where the beginning of the stop closure is not always clear. However, the aspiration period for plosives was included not as a part of a following vowel but as a part of the plosive.

## 7.4 Results and discussion

### 7.4.1 Durational ratio between moraic consonants and CV morae

First of all, the durational ratios between devoiced morae and corresponding CV morae were calculated. Two sets of measurements were taken for each devoicing site. All the ratios were averaged and the standard deviations were calculated. The results are listed in Table 7.1.

**Table 7.1** Average ratio by preceding consonants

consonant	No. of samples	mean ratio	SD
plosives	48	85.9 %	13.05
fricatives	16	88.0 %	14.96
affricates	26	77.7 %	12.42
<b>TOTAL</b>	<b>90</b>	<b>83.93 %</b>	<b>13.97</b>

Statistical analysis by T-test (two-tailed) found that there was a significant difference in the durational ratios between /CV/ morae and devoiced morae between the two types of vowels /i/ and /u/ [ $t(87)=4.21$ ,  $p < .001$ ]. Analysis by a one-way ANOVA showed that the difference in the durational ratio between /CV/ morae and devoiced morae among the three types of preceding consonants was also significant [ $F(2, 87) = 4.002$ ,  $p < .025$ ]. It seems that vowel devoicing involves not only the change of glottal state, but also durational reduction.

However, durational ratios were not as predicted in hypothesis (ii). The mean ratio of the duration between devoiced morae and their voiced counterparts for the preceding fricatives was the highest (88.0 %) among the three types of preceding consonants, and the preceding affricates was the lowest (77.7 %). If the type of preceding consonant is crucial to the duration of morae because of their physiological features, preceding fricatives and affricates should not show any difference because a preceding affricate in fact means the vowel is preceded by a



fricative. Although there was a significant difference between the types of preceding consonants, there appeared to be other reasons which caused the durational difference rather than the types of the preceding consonants.

There were 18 out of 90 cases (20 %) where devoiced morae were longer than /CV/ morae: devoiced morae preceded by plosive /k/ 8 out of 48 cases (16.7 %), preceded by affricates [tʃ] and [ts] 8 out of 16 cases (50 %), and preceded by fricative /h/ ([ç] and [ɸ]) 2 out of 26 cases (7.7 %).

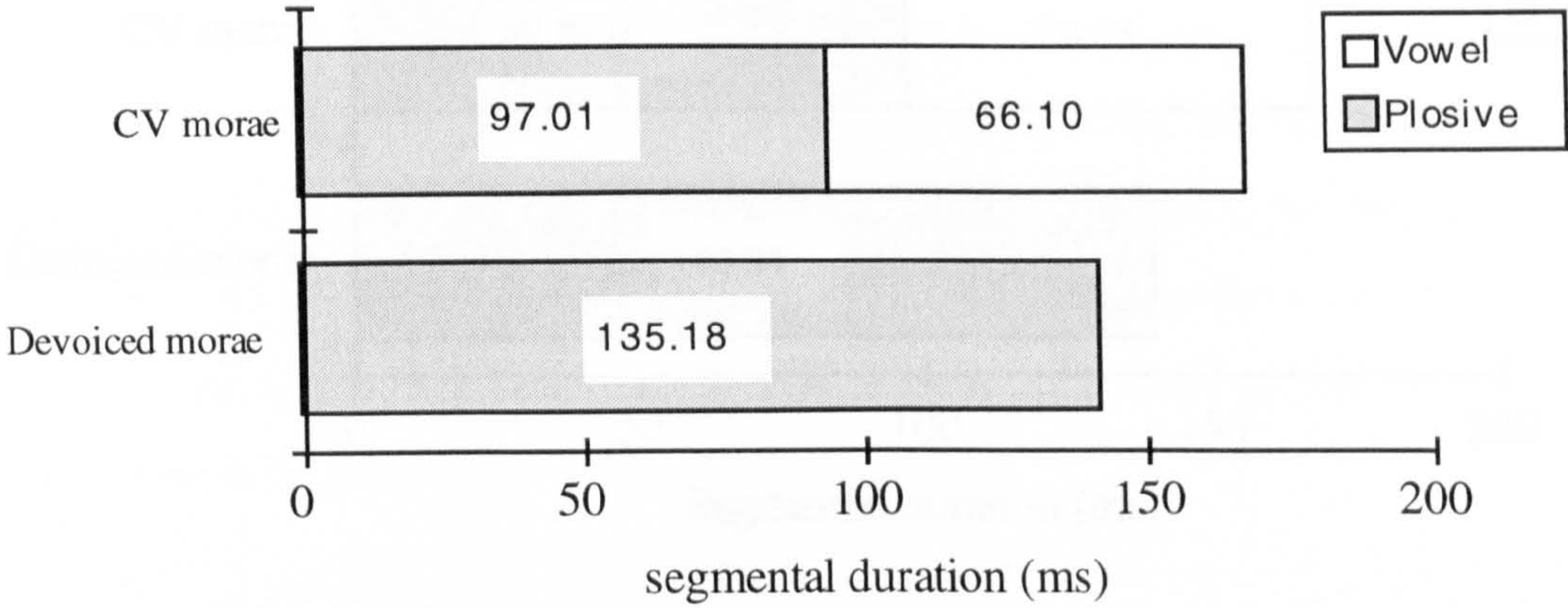
Figures 7.2a, 7.2b and 7.2c show the mean durations of devoiced morae and the mean duration of consonants and vowels in CV morae by the type of preceding consonants, and Figure 7.3 shows the mean duration of devoiced morae and the mean duration of consonants and vowels in CV morae averaged by all types of preceding consonants.

The average ratio of devoiced morae against CV counterparts was 83.93% (SD 13.97) which is much higher than the ratios obtained by Beckman. Beckman's results showed a much shorter duration of devoiced morae. In fact, they were not much longer than the consonants in CV morae. Beckman's data showed only 54% of devoiced morae (46% of stops and 61% of fricatives and an affricate) were longer than their prevocalic counterparts. In her data, there were only 7% of cases where devoiced morae were longer than CV morae compared to 14.4% in my data. Possible reasons for the difference between my results and those of the Beckman's study will be discussed in Section 7.4.4.

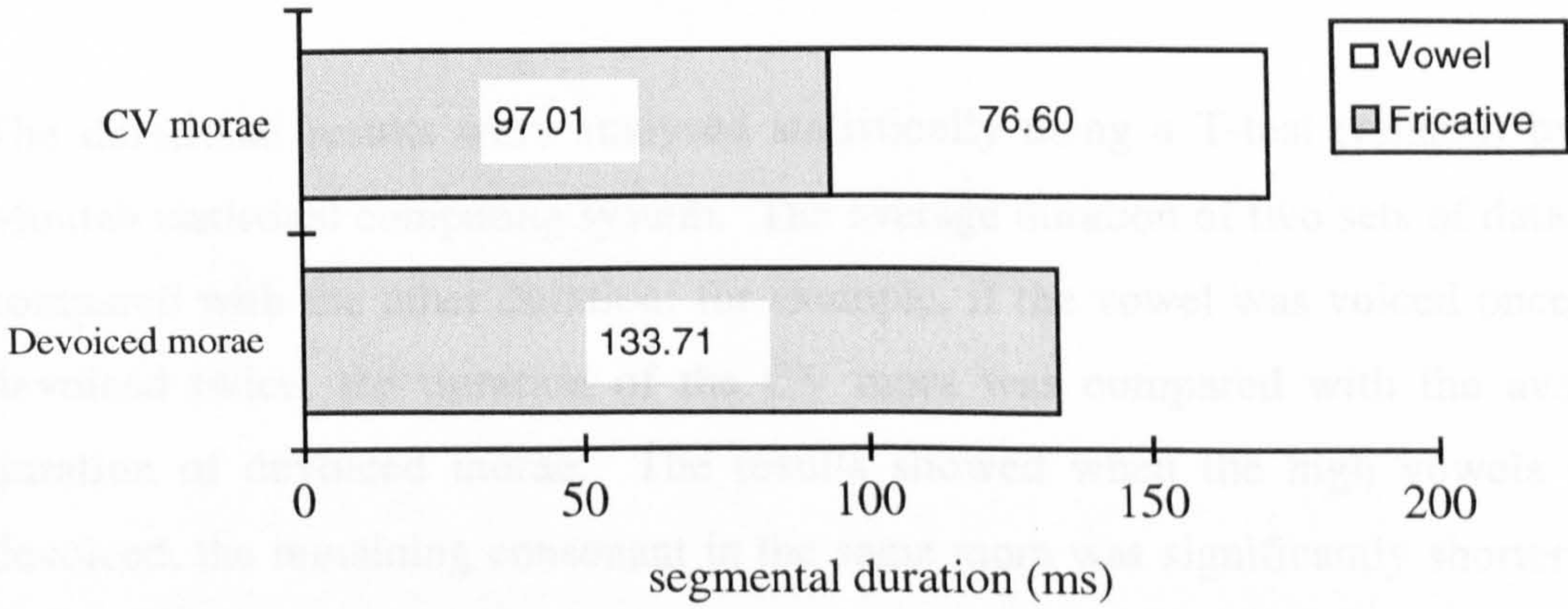


**Figures 7.2a, 7.2b and 7.2c**  
**Durational difference between devoiced morae and consonants and vowels in CV morae by type of consonants**

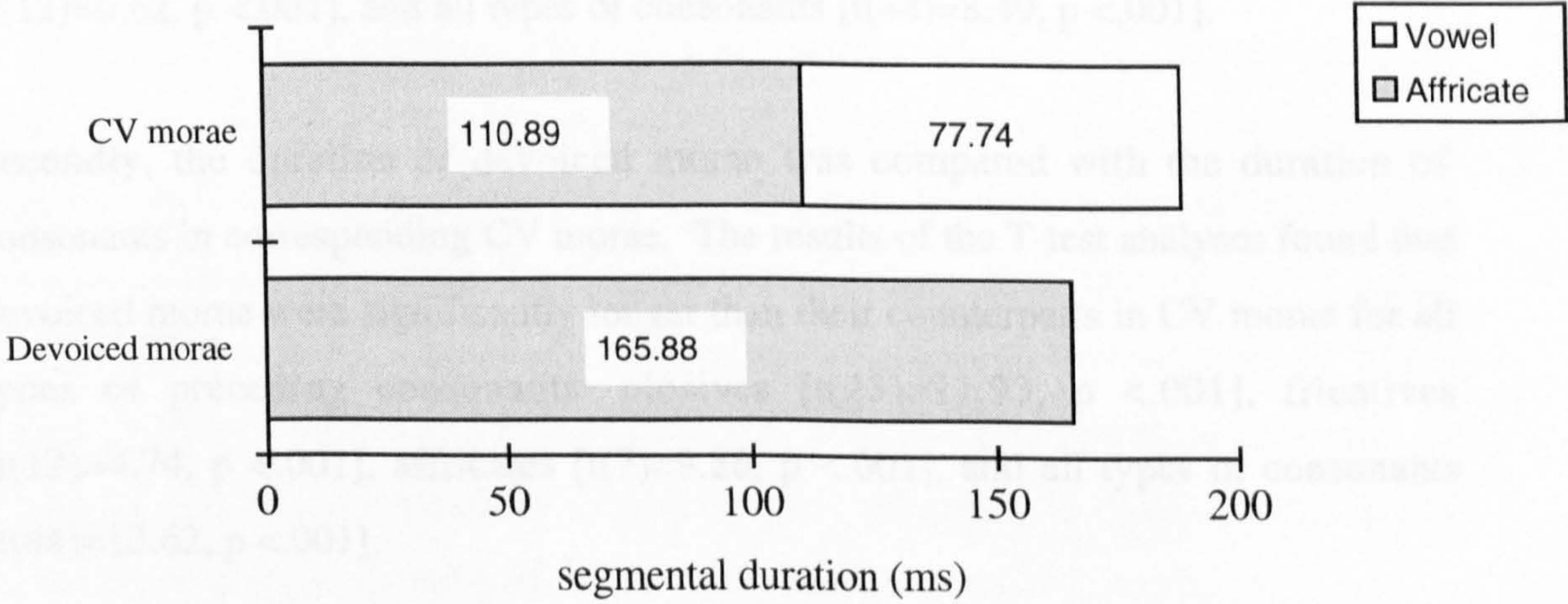
**(7.2a) Plosives**



**(7.2b) Fricatives**

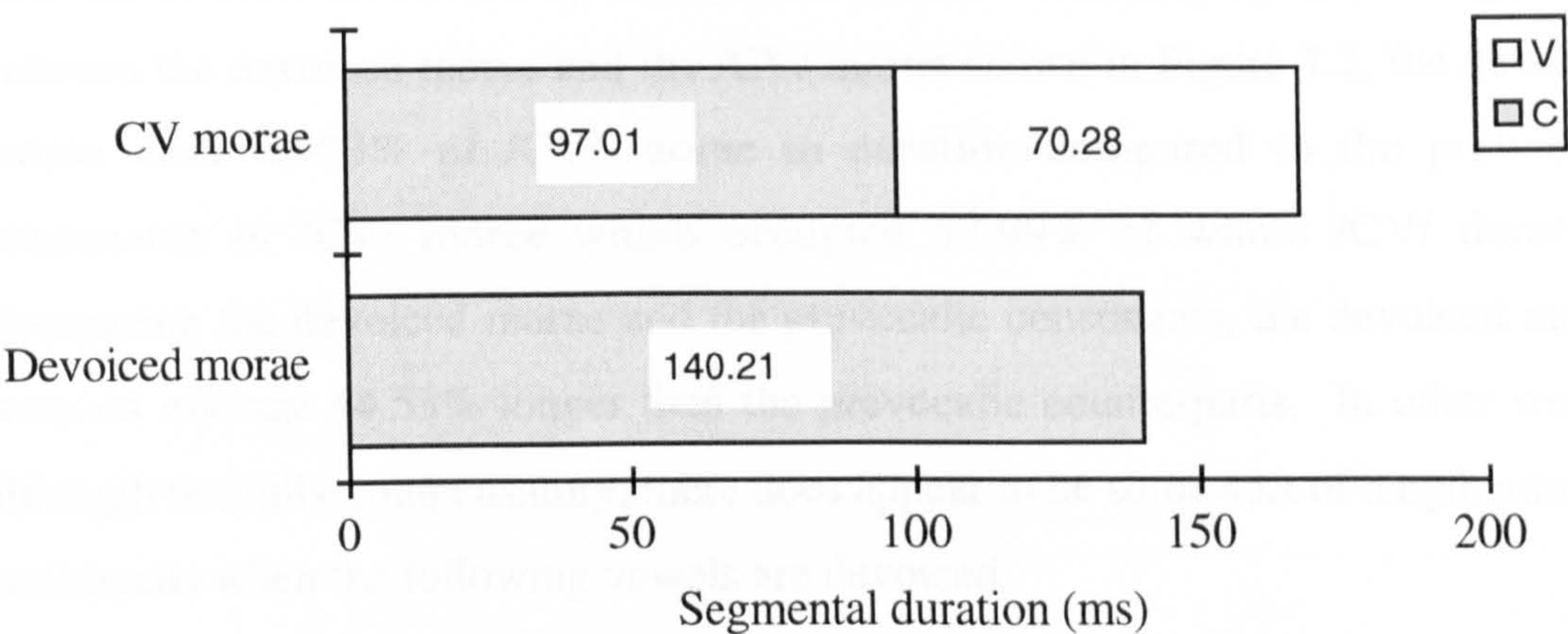


**(7.2c) Affricates**





**Figure 7.3** Average durational difference between devoiced morae and consonants and vowels in CV morae of all types of consonants



**7.4.2 Analysis by T-test**

The durational results were analysed statistically using a T-test (related) by the Minitab statistical computing system. The average duration of two sets of data was compared with the other duration: for example, if the vowel was voiced once and devoiced twice, the duration of the CV mora was compared with the average duration of devoiced morae. The results showed when the high vowels were devoiced, the remaining consonant in the same mora was significantly shorter than the equivalent CV mora, regardless of the type of devoiced vowels or preceding consonants: plosives [t(23)=5.78, p <.001], fricatives [t(7)=2.62, p <.025], affricates [t(12)=6.62, p <.001], and all types of consonants [t(44)=8.49, p <.001].

Secondly, the duration of devoiced morae was compared with the duration of consonants in corresponding CV morae. The results of the T-test analyses found that devoiced morae were significantly longer than their counterparts in CV morae for all types of preceding consonants: plosives [t(23)=11.93, p <.001], fricatives [t(12)=4.74, p <.001], affricates [t(7)=9.26, p <.001], and all types of consonants [t(44)=13.62, p <.001].



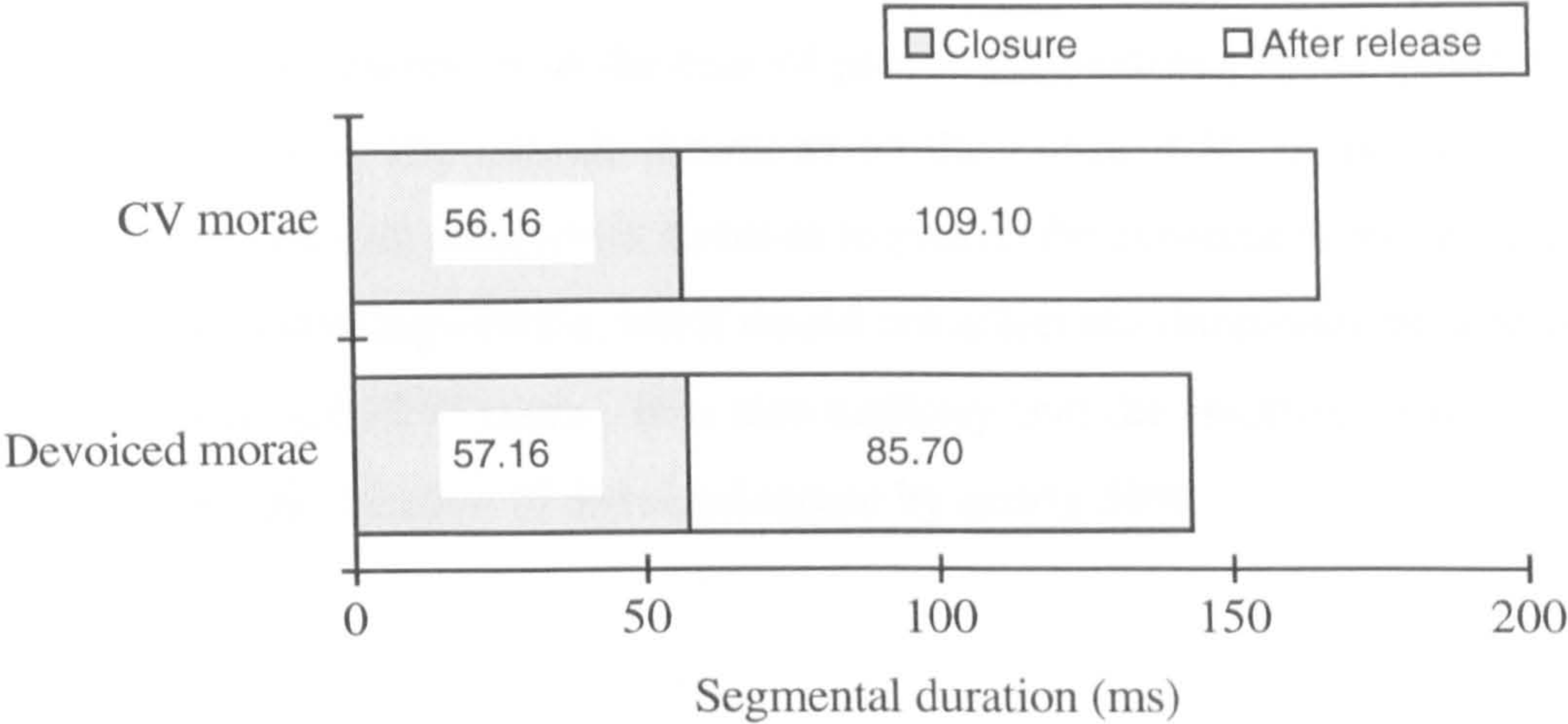
The T-test results suggested that the devoiced morae were significantly shorter than the equivalent /CV/ morae, but at the same time they were significantly longer than the consonants in corresponding /CV/ morae. Considering the average ratio between the devoiced morae and the /CV/ morae shown in Figure 7.2, the devoiced morae were 83.93% of /CV/ morae in duration compared to the prevocalic consonants in /CV/ morae which occupied 57.99% of whole /CV/ duration. Comparing the devoiced morae and the prevocalic consonants, the devoiced morae were on average 44.53% longer than the prevocalic counterparts. In other words, although not fully compensatory; there does appear to be some sort of lengthening of consonants when the following vowels are devoiced.

Thirdly, the closure durations of plosives and the plosive part of affricates in devoiced morae were compared with the closure durations of prevocalic plosives and the plosive part of prevocalic affricates in /CV/ morae. As shown in Figure 7.4, the average closure duration of plosives and affricates in CV morae was 56.16 ms (SD = 22.70), and that in devoiced morae was 57.16 ms (SD = 24.22). The T-test result (two-tailed) showed that the difference in the durations was not significant [ $t(31) = -0.51$ , n.s.].

On the other hand, the durations of the section after the release of the stop closure and fricative part of affricates in devoiced morae were very different from the total duration of this fricative part and the following vowel. The average duration of plosive and affricates in devoiced morae excluding closure duration (i.e. after release of stop closure) was 85.70 ms (SD = 24.36), whereas that of CV morae excluding closure duration of the consonants was 109.10 ms (SD = 25.16). The statistical analysis by T-test (one-tailed) found the duration of plosive and affricates after release in devoiced morae was significantly longer than that of /CV/ morae [ $t(31) = 7.12$ ,  $p < .005$ ]. The results suggested that although the durations of whole morae differed significantly depending on whether the vowel was voiced or devoiced, the closure durations of plosives and affricates in devoiced morae and those of non-moraic counterparts did not show a significant difference.



**Figure 7.4** Average closure duration and the duration after release of plosives and plosive part of affricates in CV morae and devoiced morae



**7.4.3 Problems of segmentation**

Before the durational differences will be discussed, some problems related to segmentation should be mentioned. As Beckman (1982 : 119-120) mentioned, there are some discrepancies in segmentation which may cause a systematic error in durational measurements. There is at least a period which is marginal between two segments. For instance, when a vowel immediately follows a voiceless fricative or affricate, the periodic waveform of the vowel actually starts before the frication of the preceding consonant completely ceases. However, the beginning of the vowel is generally determined to be the point where a periodic waveform starts in both the present study and Beckman’s study. Similarly in cases of voiceless plosives followed by a vowel, the beginning of the vowel is determined to be at the point where the periodic waveform of the vowel begins, regardless of whether or not the aspiration of the preceding voiceless plosive has actually ceased; this aspiration-like period, which occurs marginally between a voiceless plosive and a following vowel, sometimes appears on spectrograms as breathy resonances in higher frequencies above the lower vowel formants. On the other hand, when the vowel is devoiced, the end point of the consonant is determined as the point where the breathy resonance actually ceases.



Beckman's comments on the segmentational problems are correct. However, even if the ends of prevocalic fricative consonants were determined by the point where the frication ceases, or in the case of prevocalic plosives, by the point where the breathiness dies, the overall durations of the entire /CV/ morae would not change. This might add some extra duration to prevocalic consonants and reduce the duration of the following vowels, but it would not affect the durational ratio between devoiced morae and /CV/ mora. It is also unlikely that the frication segmentation would lengthen the duration of devoiced morae by nearly 50%.

#### **7.4.4 The duration of voiced high vowels in devoicing environments**

The experimental results showed an average durational ratio of 83.93% between devoiced morae and their CV counterparts. The ratio was much higher than the ratio obtained by Beckman (1982) (see Section 7.4.1). The T-test results showed that the devoiced morae were significantly longer than the consonants in CV morae but shorter than the whole CV morae. In other words, when a vowel was devoiced, the duration of the devoiced mora was lengthened to some extent, but not compensating fully for the lack of the voiced vowel. Does this mean that there is a partial durational adjustment of preceding consonants? Or that the alleged phonetic reality of equal duration of morae is sacrificed? Is the lengthening of the moraic consonants really durational compensation for the devoiced vowels?

When a vowel was fully voiced (Chapter 5, Figure 5.4a), periodic waves were usually repeated at least ten times on the waveform. When a vowel was partially voiced, periodic waves also occurred but repeated between three and nine times (Chapter 5, Figure 5.4b). It should be noted that when a vowel was partially voiced, it did not have clear striations on the spectrograms (Chapter 5, Figures 5.5a and 5.5b). Moreover, it was very obvious that partially voiced vowels had much shorter durations than their fully voiced counterparts: periodic waves repeated much fewer times for partially voiced vowels than fully voiced vowels. The durational data for



CV morae in this chapter included the partially voiced short vowels. This might have brought the average duration down.

Furthermore, the fully voiced vowels in the devoicing sites may also very often be shorter than vowels in non-devoicing sites. While admitting that (1) the high vowels have inherently shorter durations, and (2) the high vowels in devoicing sites are surrounded by voiceless consonants which are inherently longer than their voiced counterparts, the vowel duration may be slightly adjusted and shortened, if there is a mora-level durational adjustment between neighbouring segments.

The results in Chapter 5 showed when there was only one devoicing site in a word, the vowel was almost always devoiced. Therefore, most of the durational data for this experiment were from words with more than one devoicing site. If there is more than one devoiceable vowel in a word or breath group, it is still acceptable for the devoiceable vowels to undergo a devoicing process, but certain constraints block some devoiceable vowels from undergoing the devoicing process. The devoiceable vowels which remain voiced, however, do not often have the same duration as vowels in non-devoicing sites. Their durations tend to be shortened; voiced devoiceable vowels very often had fewer periodic waves than the same vowels in the non-devoicing environment pronounced by the same speaker. This was possibly the reason why the durational ratios of devoiced morae to /CV/ morae were very high. This means that the high durational ratio of devoiced morae might not be due to the compensatory lengthening of the consonants to cover the duration of the lacking vowel, but rather due to the durational reduction of /CV/ morae in typical vowel devoicing environments.

However, the other durational comparison, between devoiced morae and prevocalic consonants in /CV/ morae, also showed significantly longer durations for devoiced morae. If it were the high vowels in /CV/ morae whose durations were reduced, there would be no reason why the devoiced morae were significantly longer than the consonants in /CV/ morae.

To explain the difference between our results and those of Beckman (1982) which showed a much smaller ratio between devoiced morae and the /CV/ morae, and also very little difference in the ratio between devoiced morae and non-moraic consonants in the /CV/ morae, we may turn to differences in the devoicing environments.

Beckman used 23 test words for her experiment, among which 13 words had only one devoicing site. The other words had vowel devoicing possibilities in two consecutive morae. Her five subjects varied voicing of high vowels in 27 instances. She did not mention consecutive devoicing cases. Probably her subjects did not devoice the vowels in two consecutive morae. The data for my experiment were, on the other hand, taken mainly from words with more than one devoicing site. The subjects very often devoiced vowels in two consecutive morae, and voicing of vowels varied to a large extent (see Chapter 5, Section 5.4.5).

As discussed earlier in this chapter, the phonetic reality of mora-timing and durational adjustment within a mora are still controversial. It does not appear to be true that each mora has exactly the same duration. However, it seems to be true that Japanese has a tendency to maintain the duration of words or breath groups proportional to the number of morae, although this is not perfect (Port et al., 1980, Sugito, 1989, Y. Sato, 1993, Han, 1994, etc.). There is also some sort of durational adjustment between the members of a CV pair in the same mora and the members of a V-C sequence across a mora boundary (Higuchi and Fujisaki, 1980, Sagisaka and Tohkura, 1981).

In the cases of consecutive vowel devoicing, if there is very little durational adjustment of the preceding consonants when the following vowels are devoiced, it creates a succession of shorter morae. My earlier results showed that at a normal speaking tempo, vowel devoicing in consecutive morae was possible but not in more than two morae (see Chapter 5). This might possibly be partly because the duration of devoiced morae is generally shorter than CV morae and consecutive devoicing creates a series of short morae: Japanese may not tolerate a succession of shorter

morae if the duration is too short to be adjusted for in other segments or morae in the same word or breath group. Moreover, even when devoiced vowels remained voiced, the duration of voiced devoiced vowels tended to be shorter (partially voiced) than the same vowels in the non-devoicing environment. Thus, it seems plausible that in the consecutive vowel devoicing environments, the remaining consonants in devoiced morae may be lengthened in order to maintain the duration of an entire word.

#### **7.4.5 Durational relationship between a consonant and a vowel in the same mora**

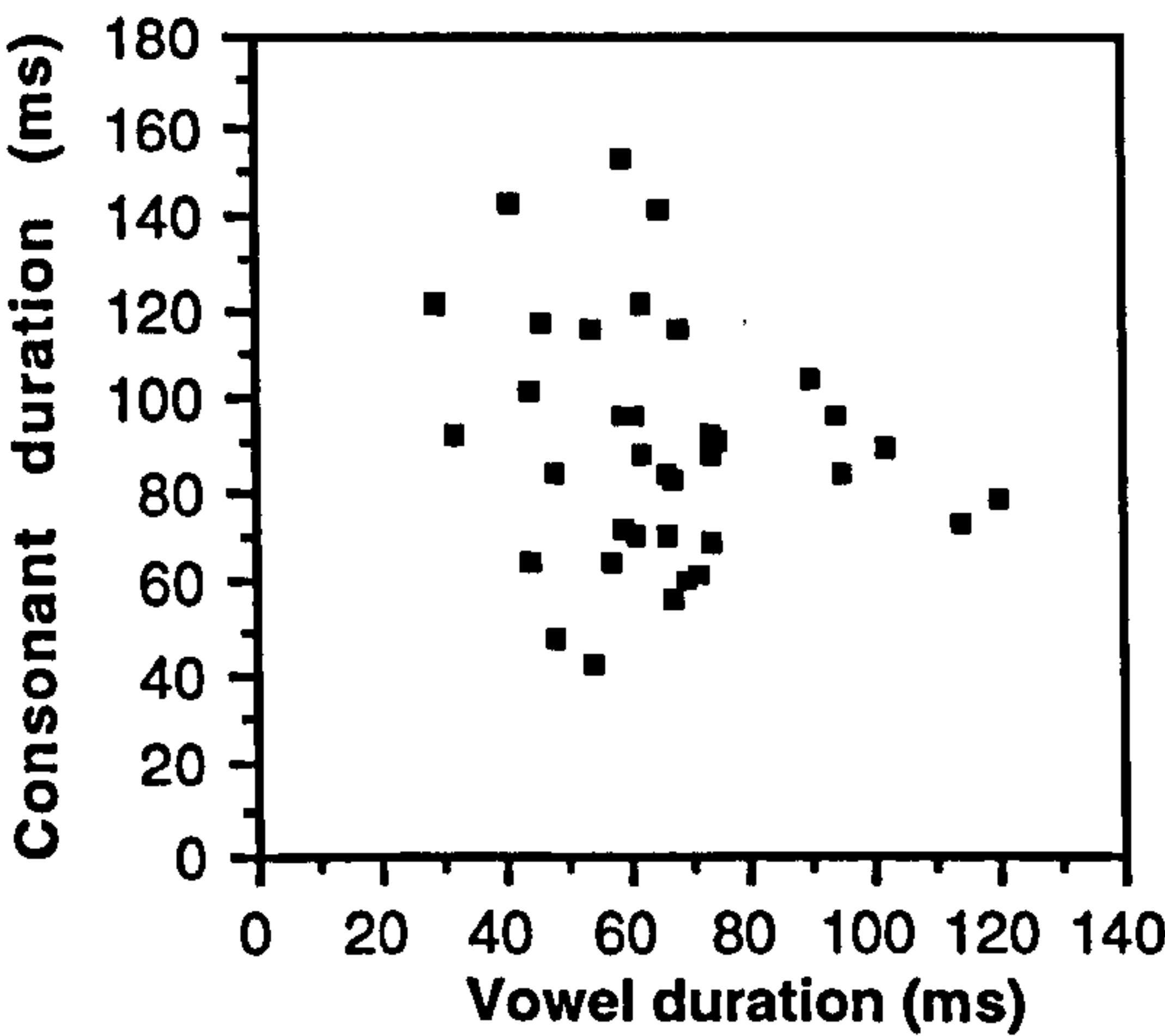
Although devoiced morae were significantly shorter than /CV/ morae (83.93 %), they were significantly longer than the non-moraic consonants in /CV/ morae. This durational difference might have been caused not only by the partial durational adjustment of the remaining consonants in devoiced morae, but also by the durational reduction of the voiced vowels in /CV/ morae; although devoiced morae were shorter than /CV/ morae, the ratio between devoiced morae and /CV/ morae was quite high.

On the other hand, if there is durational adjustment within a mora in order to keep the duration of each mora similar, the shorter duration of partially voiced vowels should not affect the duration of whole /CV/ morae, as the duration of the preceding consonants would be lengthened. To examine whether durational adjustment within a mora really operates in devoiced morae, the relationship between consonant and vowel durations within the same mora was studied.

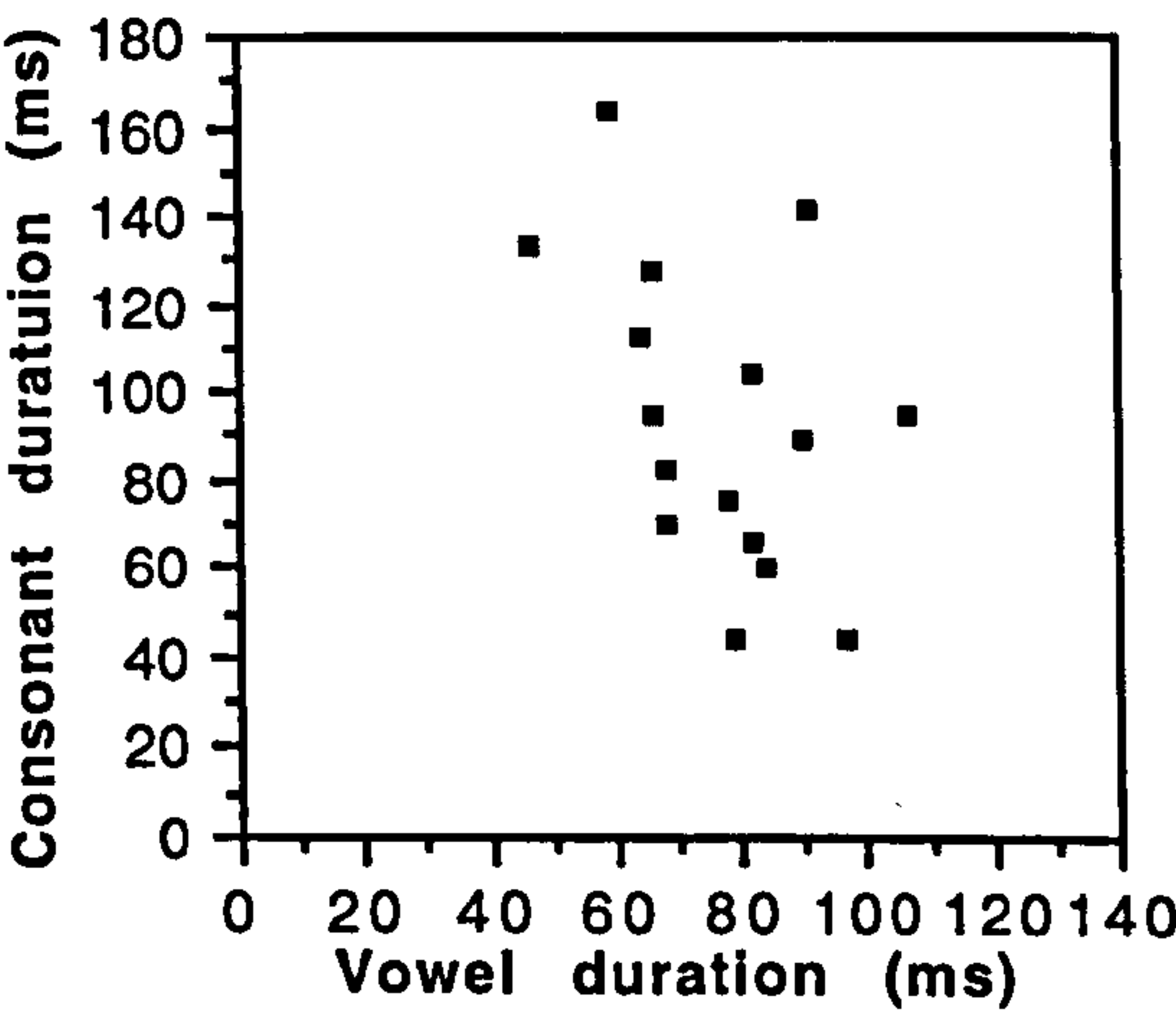


**Figures 7.5a, 7.5b and 7.5c**  
**The durational relationship between consonants and vowels within a mora by the type of consonants**

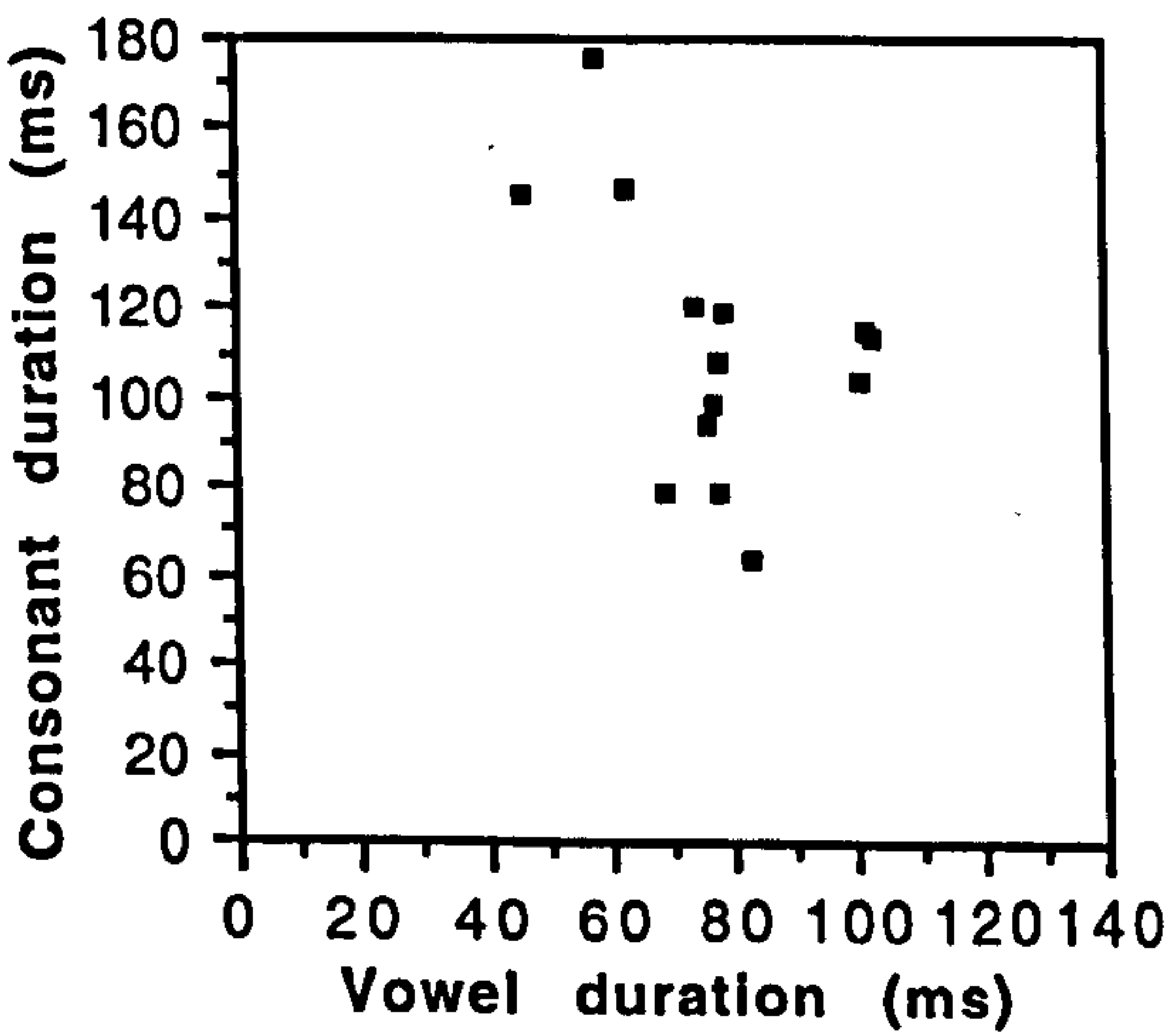
**(7.5a) Plosives**



**(7.5b) Fricatives**

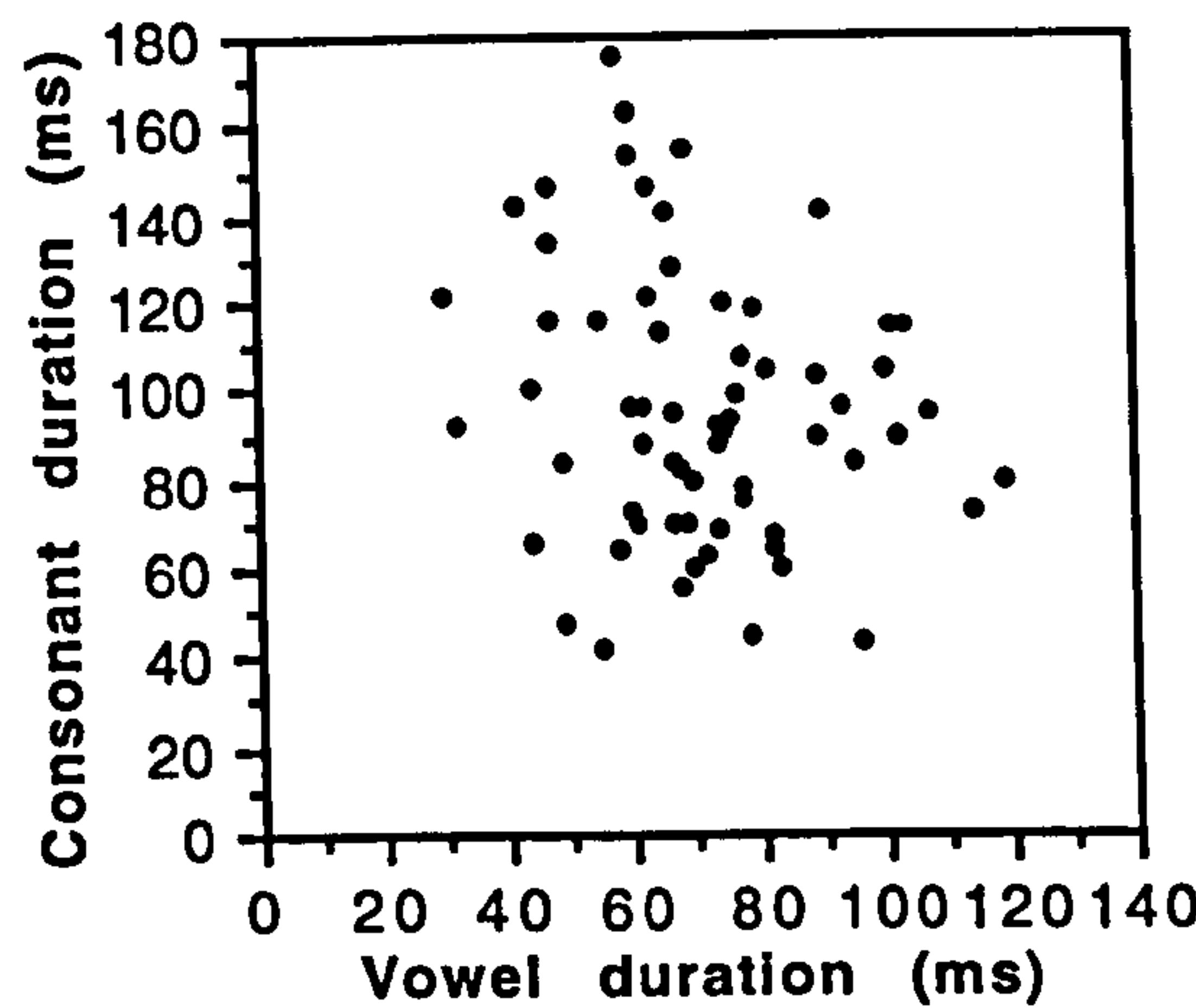


**(7.5c) Affricates**



The duration of each vowel in a /CV/ mora was plotted against the duration of the consonant in the same mora. The results were shown in Figures 7.5a, 7.5b, 7.5c and 7.6.

**Figure 7.6** The durational relationship between consonants and vowels in the same morae for all types of consonants



Although generally there were tendencies of negative correlation for all types of preceding consonants, the statistical analysis one-tailed correlation found that the results were not significant except for preceding fricatives as shown in Table 7.2.

**Table 7.2** The results of correlation between consonant duration and vowel duration by types of preceding consonants

Type of consonant	Correlation result
plosives	$r(35) = -0.196, \text{ n.s.}$
fricatives	$r(14) = -0.442, p < .05$
affricates	$r(11) = -0.445, \text{ n.s.}$
all consonants	$r(64) = -0.196, \text{ n.s.}$

The results suggested that there seemed to be some durational adjustment between neighbouring segments, but the adjustment did not operate within morae in order to maintain the duration of each mora constant. Therefore, the cause proposed earlier in Section 7.4.4 of the higher ratio between devoiced morae compared with the /CV/ morae, that it might have resulted from the reduction of voiced devoiceable vowels, might be plausible.

There appeared to be no durational adjustment between consonant and vowel within a mora in the devoicing environment. However, a durational adjustment may still be in operation at the level of larger units such as a word rather than the mora. If there is a word level durational adjustment, the greater the number of devoiced vowels in a word, the more durational adjustment would operate in each devoiced mora as well as in the other parts of the word, so that the overall word duration would remain constant, depending on the number of morae in a word.

In the next section, the degree of durational adjustment of devoiced morae and the number of devoiced vowels in a word will be examined.

#### **7.4.6 The duration of devoiced morae and the number of devoiced vowels in a word**

The durations of devoiced morae and their voiced counterparts were examined with relation to the number of morae in a word to test hypothesis (iii) in Section 7.2. The duration of each devoiced mora was compared with its voiced counterpart, and the durational ratio (“duration of devoiced mora” divided by “duration of voiced /CV/ mora”) was calculated, obtaining 90 ratios (2 comparisons for 45 triplets). The ratios were classified according to the number of morae in a word in which devoicing occurred and the number of devoiced vowels in the same word. The prediction is that when there are more devoiced vowels in a word, the duration of each devoiced mora will be closer to the duration of the CV mora, but when there is only one



devoiced vowel in a word, the duration of the devoiced mora will be much shorter than the CV mora. If all devoiced morae are significantly shorter than CV morae, and there are several shorter morae in a word, shorter durations of devoiced morae may not be fully compensated for at the word level. The durational ratios by the number of devoiced vowels in a word are listed in Table 7.3.

The results showed that the ratio between a devoiced mora and a /CV/ mora was the highest when there were 3 devoiced vowels in a word (90.86 %) as predicted, but the ratio in words with 1 devoiced vowel was higher than the ratio in words with 2 devoiced vowels in a word. A statistical analysis by a two-way ANOVA taking durational ratio between a voiced morae and a /CV/ mora as the dependent variable and the number of devoiced vowels and the number of morae in a word as factors showed that the only significant effect on the duration was the number of morae in a word [ $F(3,81)=5.723$ ,  $p < .01$ ]. The effect of the other factor “the number of devoiced vowels in the same word” was not found to be significant [ $F(2,81)=2.304$ , n.s.], and there was no significant interaction between the number of devoiced vowels and the number of morae in a word [ $F(3,81)=1.702$ , n.s.].

**Table 7.3    The durational ratios between voiced and devoiced morae by the number of devoiced vowels in a word**

	1 devoiced vowel	2 devoiced vowels	3 devoiced vowels
Mean ratio (SD)	84.49% (13.94)	81.70 % (14.59)	90.86 % (10.48)
No. of examples	62	25	3

Secondly, the durational ratios were analysed in relation to the number of morae in a word. The statistical results show that the number of devoiced vowels in a word has no influence on the durational ratios between devoiced morae and CV morae as shown in Table 7.4.

The results suggested that the number of devoiced morae in a word did not affect the duration of devoiced morae. The experimental hypothesis was found untenable that

the greater the number of devoiced vowels in a word, the closer the duration of the devoiced mora to the duration of CV morae.

**Table 7.4** The results of One-way ANOVA (related) by the number of morae in a word as a factor

No. of morae in a word	Statistical result
4 mora words	F(1,18)=.293, n.s.
5 mora words	F(2,29)=1.250, n.s.
6 mora words	F(1,12)=3.775, n.s.
7 mora words	F(1,28)=.253, n.s.

The results in Sections 7.4.5 and 7.4.6 imply that a shorter duration of the devoiced morae might not be adjusted for at the word level, so that the whole duration of a word would be consistently dependent upon the number of morae in a word. If the duration is really adjusted at the word-level as suggested by Port et al. (1987), the temporal adjustment works beyond the segmental or moraic level and is not significantly affected by durational irregularity of smaller units than a word.

#### 7.4.7 Reality of mora-timing

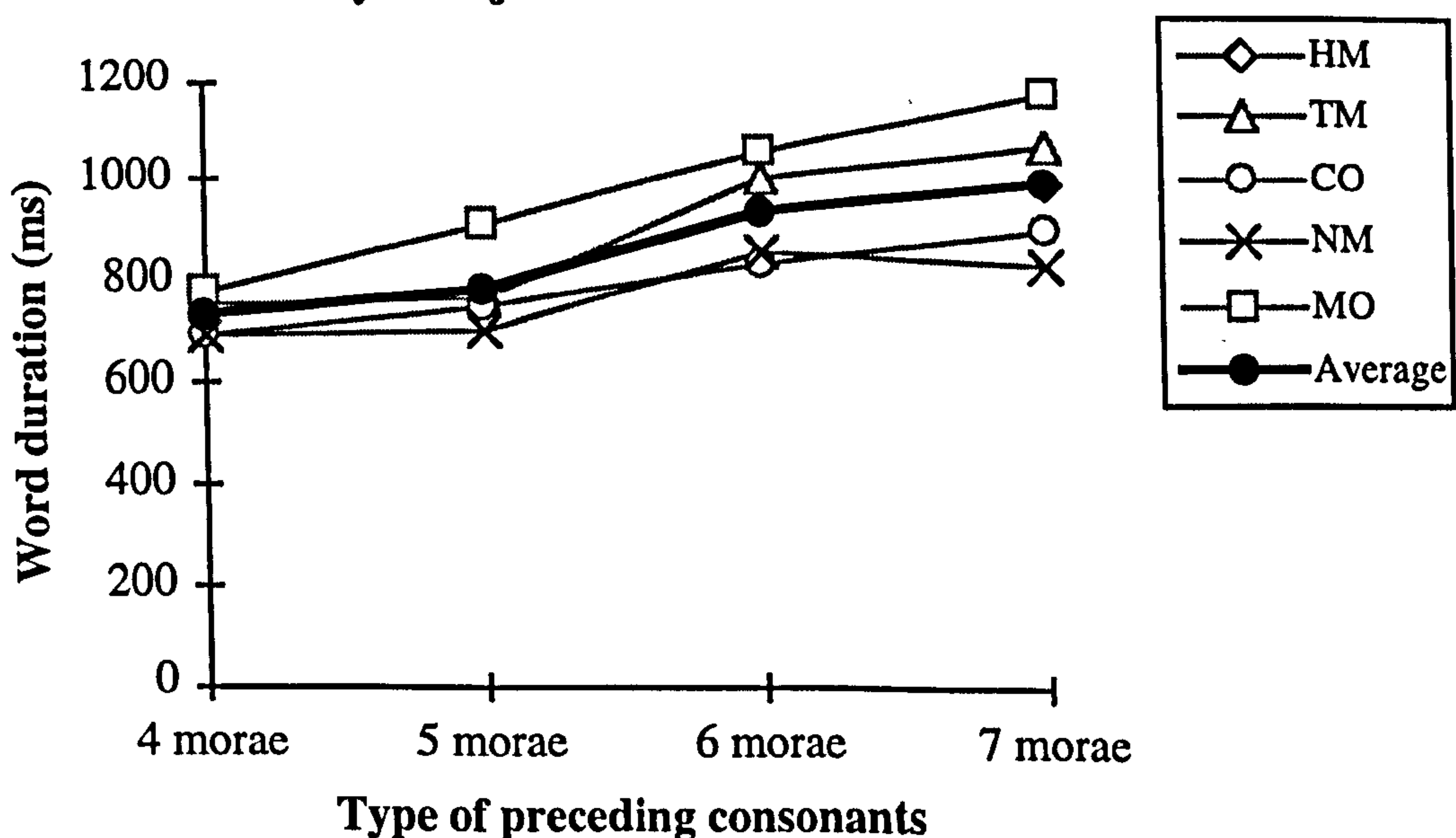
The above results showing shorter durations of devoiced morae mean that the durations of voiced morae would also have to be adjusted to agree with the moraic duration theory of Port et al. (1987). Therefore, the durations of whole words were measured and examined with relation to (i) the number of morae in a word, and (ii) the number of devoiced morae in a word.

Since the test words were pronounced in citation form, the last mora of a word was usually lengthened, and the initial mora of a word was usually shortened (Sagisaka and Tohkura, 1984 and Takeda et al. , 1989). Moreover, the duration of the word initial stop closure and the beginning of fricatives were not always clear. In these

cases, the closure duration and fricative duration were excluded from the total word durations. Therefore, the measurements were not the correct measurements of whole words, but they might show a general tendency. A comfortable speaking rate varied from subject to subject. The number of analysed examples for each mora for each subject differed enormously. For example, Subject HM did not show any voicing variation in 4 and 6 mora words. Only Subject MO showed a steady increase of duration according the number of morae in a word. However, as shown in Figure 7.6, there was a tendency in general for the duration of a whole word to be lengthened as the number of morae in the word increased.

The number of morae in a word varied from 4 to 7, and the number of devoiced vowels in a word varied from 0 to 3. As listed in Table 7.5, statistical analysis by ANOVA (3-way) found that the word duration was significantly influenced by the subject [ $F(4,81)=34.714$ ,  $p < .0001$ ], the number of morae in a word [ $F(3,81)=95.570$ ,  $p < .0001$ ], and the interaction between the subject and the number of morae in a word. However, the effects of other factors and interactions between these factors were not found significant.

**Figure 7.6** The duration of whole words by the number of morae in a word and by 5 subjects





**Table 7.5    The results of three-way ANOVA by the number of devoiced vowels in a word, the subject and the number of morae in a word as factors**

Factor	df	F-value	p
(a) The number of devoiced vowels in a word	3	.576	n.s.
(b) Subject	4	34.714	p <.0001
(c)The number of morae in a word	3	95.570	p <.0001
(a) x (b)	9	.643	n.s.
(a) x (c)	7	1.474	n.s.
(b) x (c)	10	3.934	p <.001
(a) x (b) x (c)	13	.796	n.s.
Residual	81		

Secondly, the word durations of each number of morae were analysed using ANOVA (2-way) by the subjects and the number of devoiced morae in a word as the factors. The results shown in Table 7.6 found that for all numbers of morae in a word, i.e. 4 to 7 morae, the effect of the number of devoiced vowels was not significant. No significant interaction was found between the two factors.

**Table 7.6    The results of 2-way ANOVA of the effects of factors (a) subject and (b) the number of devoiced vowels in a word on word duration**

No. of morae in a word	Factors		Interaction
	(a) Subject	(b) No. of devoiced vowels in a word	(a) x (b)
4	F(3,19)=7.931, p <.01	F(2,19)=4.095, n.s.	F(5,19)=1.857, n.s
5	F(4,20)=8.551, p <.001	F(3,20)=.187, n.s.	F(7,20)=.809, n.s.
6	F(3,12)=18.010, p <.001	F(3,12)=.123, n.s.	F(5,12)=.901, n.s.
7	F(4,30)=25.532, p <.001	F(2,30)=1.359, n.s.	F(5,30)=.215, n.s.

The statistical results showed that the number of devoiced vowels in a word was not an important factor for the whole duration of words. Rather, it was the number of morae in a word and individual speakers that significantly influenced the whole duration of words. This may imply that the shorter durations of devoiced morae were adjusted at a word level so that the whole duration of a word did not have to change too much as Port et al. (1987) and Han (1994) demonstrated.

## 7.5 Conclusions

Durational measurements of devoiced morae and /CV/ morae showed that the remaining consonants in the devoiced morae were significantly shorter than /CV/ morae. In other words, the proposed tendency of equalising mora duration was not tenable in devoiced morae. The moraic consonants were significantly longer than the consonants in /CV/ morae. However, the closure durations of moraic plosives and affricates, and those of prevocalic plosives and affricates were very similar. That means that the remaining consonants in devoiced morae were not lengthened in order to compensate for the lack of a vowel, but rather the vowel was underlying with the vowel portion hidden, although the underlying vowel was not the full length of its voiced counterpart.

As for mora-based durational adjustment at the word level, the experimental results showed that the number of devoiced vowels in a word did not affect either the duration of a whole word, or the ratio between a devoiced mora and a /CV/ counterpart. There was a tendency for a negative correlation between consonant durations and vowel durations within a mora, but statistically these correlations were not significant. These results imply that shorter durations of devoiced morae were not adjusted within a mora, but beyond the mora as suggested by Port et al. (1987).

The results suggested that contrary to similar processes in other languages, the Japanese vowel devoicing process does not involve durational reduction, although the duration of the devoiced mora might become shorter than the voiced CV

counterpart, when a vowel was devoiced. However, the duration of the whole word remained constant. Therefore, it is not appropriate to call the process 'vowel reduction', but probably more appropriate to call it 'vowel weakening' as discussed in Chapter 2.

In the next chapter, vowels in devoicing environments and in non-devoicing environments will be compared with particular reference to their qualities and intensities, and will be examined in relation to whether the vowel devoicing process involves vowel centralisation.



## Chapter 8

# Vowel devoicing as a weakening process in Japanese

### 8.1 Introduction

In this chapter, I will examine the proposal that Japanese vowel devoicing is a weakening process. In Chapter 2, I examined the phenomenon called vowel devoicing in various languages, and found that the typical vowels which undergo devoicing are inherently short or weak vowels, namely high vowels and schwa, and the typical environment for vowel devoicing is adjacent to a voiceless sound. It was also found that vowel devoicing occurs as part of the vowel reduction process and devoiced vowels are often actually deleted at the end of the process. Vowel reduction, devoicing and deletion are part of the overall process of vowel weakening in many languages, and occur typically in fast and casual speech. In the languages examined, stress often plays an important role in the rhythm of the language, and vowels which undergo weakening processes are typically in unstressed positions, with centralisation of vowel quality and durational reduction typically part of the overall process of vowel weakening.

Japanese vowel devoicing occurs in similar environments to other languages, namely between voiceless consonants, or between a voiceless consonant and a pause. However, Japanese vowels do not have the characteristics of obvious centralisation and durational reduction related to accentuation (Keating and Huffman, 1984). Devoicing is not a typical process just of fast or casual speech, but also occurs in

formal or slow speech (see Chapters 5 and 6). However, Japanese vowel devoicing is not a categorical process, switching abruptly between voiced and voiceless. The phonetic realisation of vowels in the devoicing environment varies from a realisation as a voiced vowel to a realisation as a completely voiceless vowel, or sometimes the vowel is even deleted (see Chapter 5, Section 5.4.1).

Partially voiced vowels generally had a shorter duration than fully voiced vowels. Periodic waves usually repeated 6 or 7 times, but often only three or four times. Fully voiced vowels, on the other hand, had a longer duration; on the speech pressure waveform, clear periodic waves were repeated for a period definitely recognisable (at least 10 times) and they showed clear dark striations on the spectrogram. Partially voiced vowels were also generally weaker than fully voiced vowels; their intensities were lower. In Chapter 5, criteria for differentiating fully and partially voiced vowels were put forward: if periodic waves repeat from 3-4 times up to 9 times, vowels are categorised as ‘partially voiced’, and if periodic waves repeat at least 10 times, vowels are categorised as ‘fully voiced’. The problem is that it is not accurate to rely on the number of periodic waves to determine the degree of voicing. Everything being equal, the lower the pitch of the voice, the fewer the number of periodic waves, i.e. lower frequency. Therefore, generally speaking, male voices tend to have fewer periodic waves. As it was not easy to draw an absolute boundary between ‘fully voiced’ and ‘partially voiced’ vowels, both types of vowels were treated as ‘voiced’ vowels in the analyses.

Figure 8.1 shows a typical example of a partially voiced vowel in a devoicing environment. The underlined vowel /u/ in the word /to.o.-ho.-ku-ti'.-ho.o/<sup>1</sup> [to:hokɯtɕiho:] ‘the Touhoku District’ is in a typical devoicing environment surrounded by voiceless consonants /k/ and /tɕ/, but it happened to be voiced in this particular utterance. Clear periodic waves can be seen on the waveform and striations on the spectrogram, but its duration is much shorter than the durations of other short vowels in the same utterance. /o:/ in the initial and final syllables is both phonemically and phonetically long; thus it is not reasonable to compare their

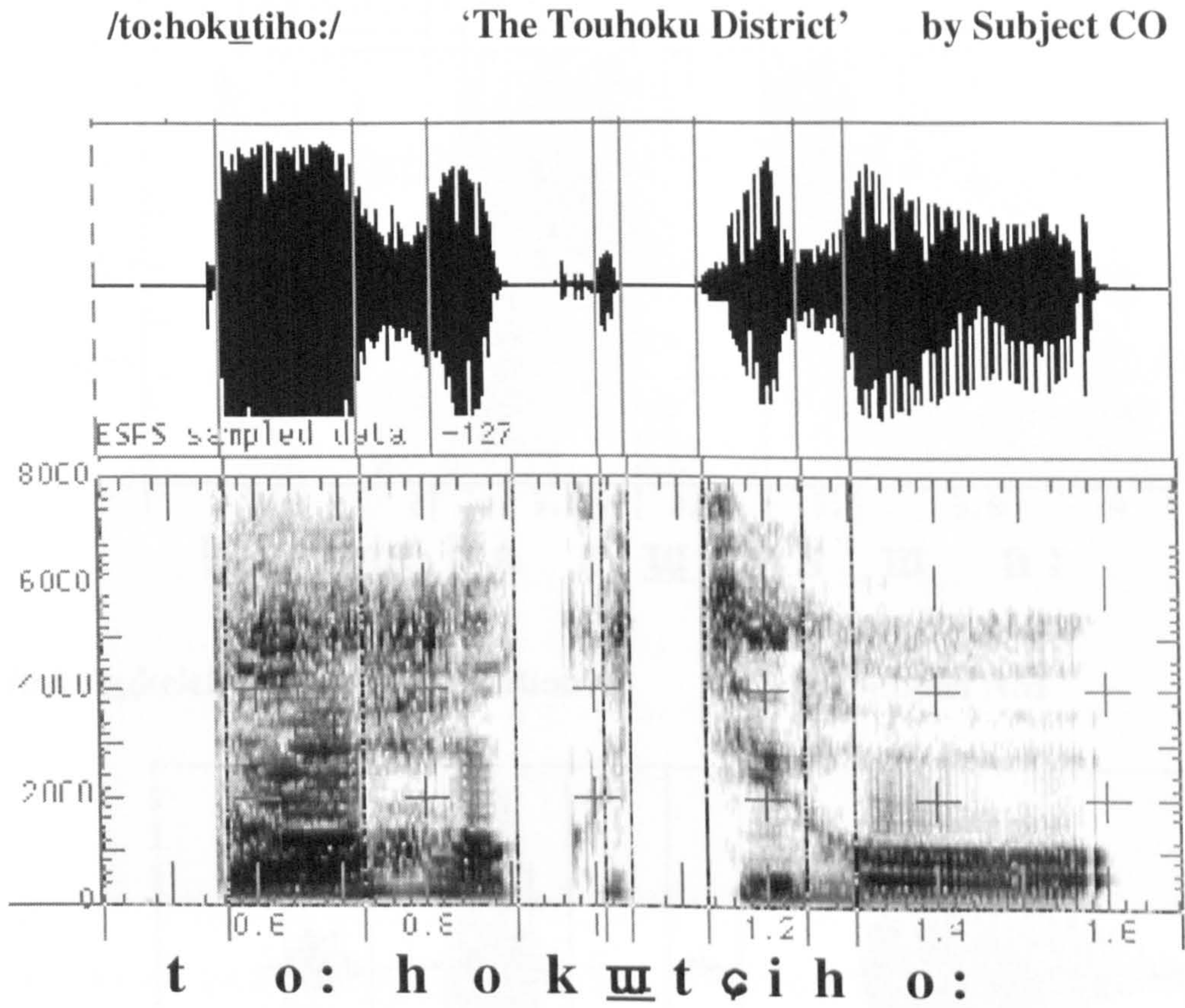
---

<sup>1</sup> /. / indicates a mora boundary, and /- / indicates a syllable boundary.



durations with the duration of the /u/. Considering the fact that high vowels are inherently shorter than non-high vowels (see Chapter 2), and that neighbouring consonants greatly influence the duration of the vowel (see Chapter 7), compared with the other phonemically short vowels /o/ and /i/ in the same utterance, the /u/ is still significantly shorter.

Figure 8.1 Example of a partially voiced vowel

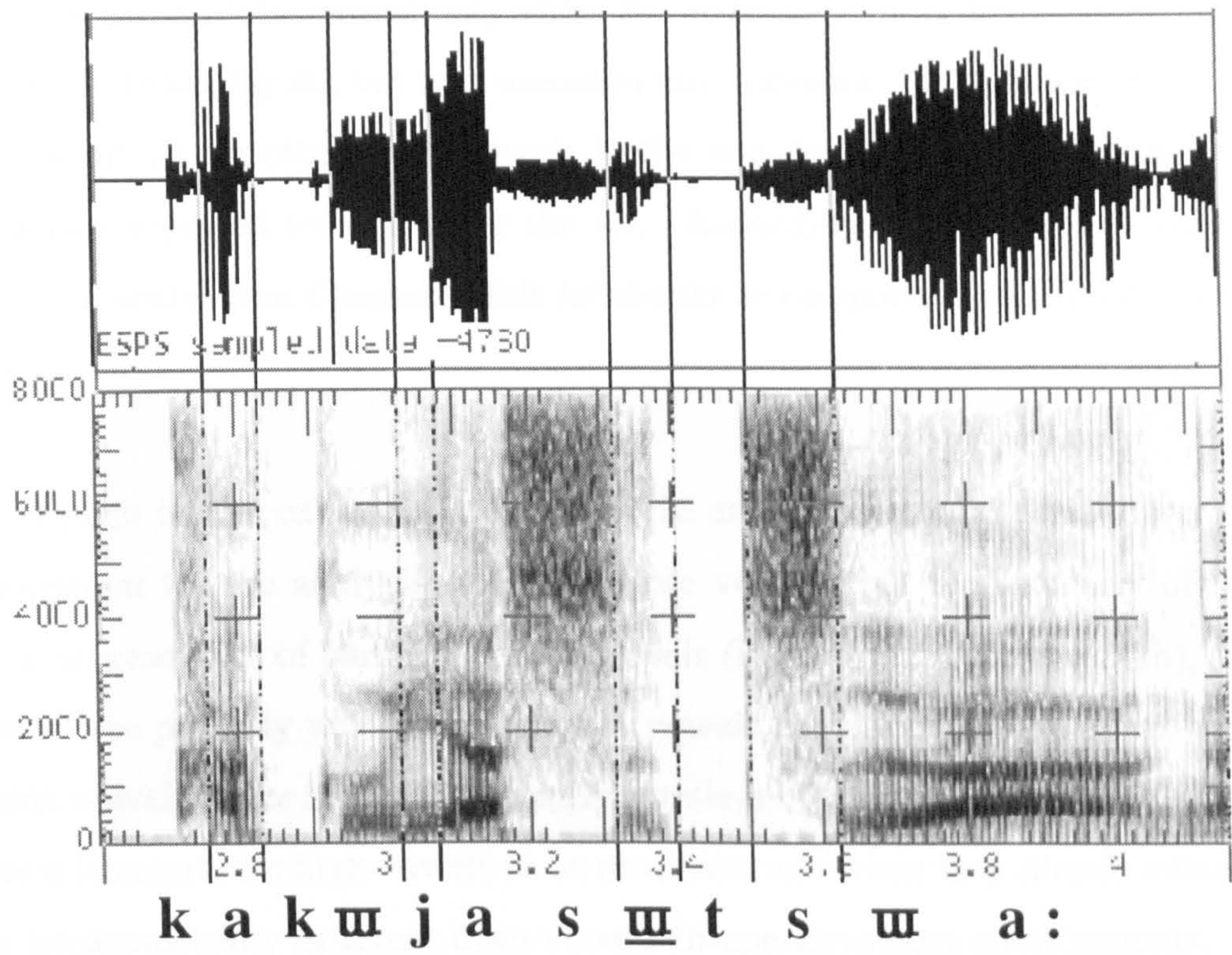


Figures 8.2a and 8.2b also present examples of the partially voiced vowel /u/. Figure 8.2a shows the waveform and spectrogram of the utterance /ka.-ku.-ja.-su.-tu.-a.a/ [kakɯjasutsɯa:] ‘cheap package holiday’, in which the underlined /u/ in the fourth mora /su/ is devoiceable surrounded by the voiceless consonants [s] and [ts], but it was voiced in this utterance. The duration of the voiced devoiceable /u/ in this utterance was not particularly shorter than those of other short vowels in the same utterance. It is difficult to see the durational difference between them by observation.



Figure 8.2 Examples of intensity weakening

(a) /kakujasutua:/ ‘cheap package holiday’ by Subject TM



(b) /takakukeiei/ ‘diversification’ by Subject NM

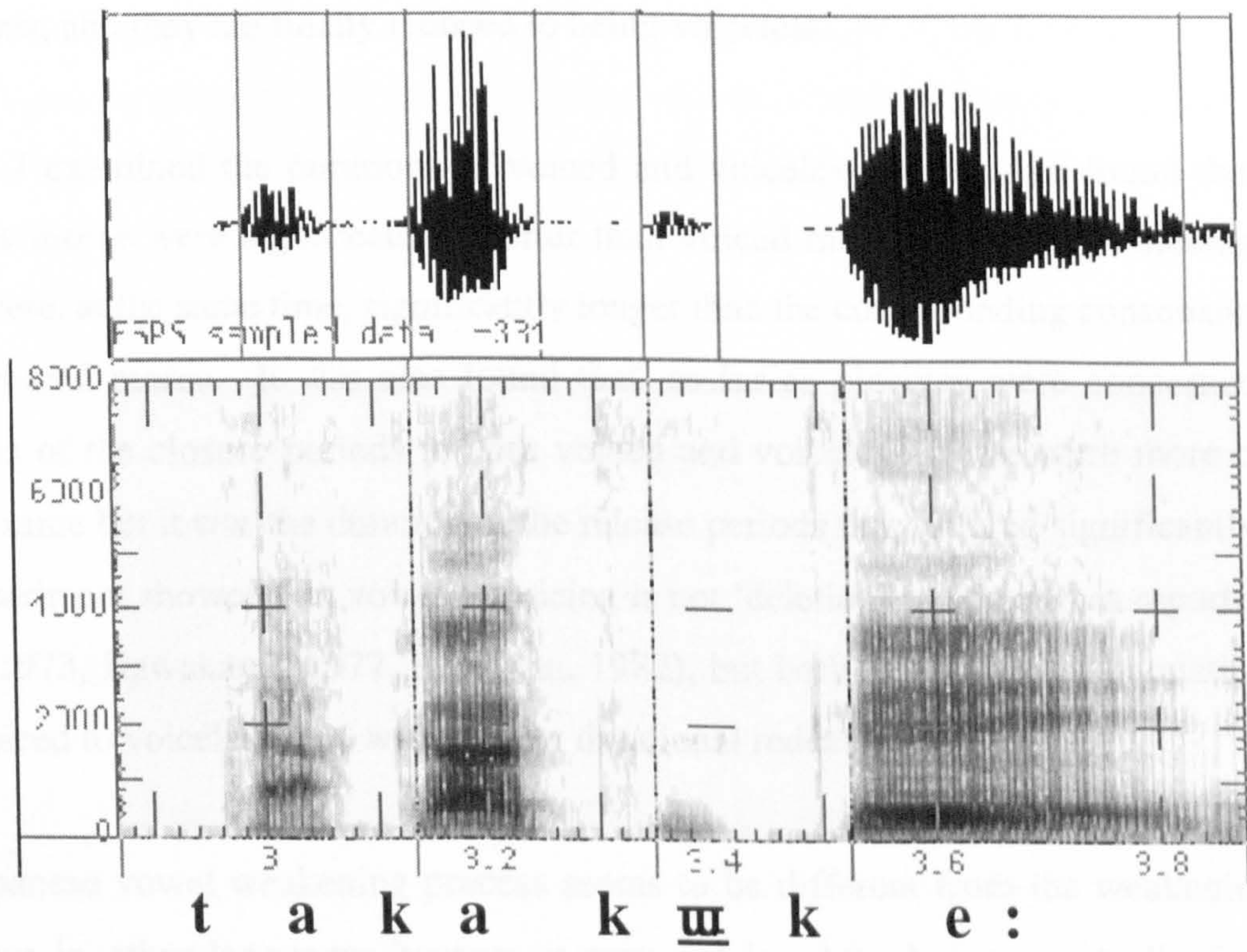




Figure 8.2b shows the waveform and spectrogram of the utterance /ta.-ka.-ku.-ke.i-e.i/ [takakuke:e:] ‘diversification’, where the underlined /u/ is devoiceable, both preceded and followed by /k/, but was voiced in this utterance. Its duration seems to be slightly shorter than other short vowels in the non-devoicing environment, but periodic waves repeated ten times for the /u/. According to the voicing criteria decided for the analyses in Chapter 5, this /u/ should be categorised as a fully voiced vowel.

The crucial point for the examples in Figures 8.2a and 8.2b is that it is not duration that is important for the analysis of devoiceable vowels, but the intensity of the vowels. In all examples of partially voiced vowels (Figures 8.1, 8.2a and 8.2b), the intensities of the partially voiced devoiceable vowels are lower than those of non-devoiceable vowels in the same utterances, regardless of their durations. This may imply that it is natural for high vowels to be devoiced, and when they remain voiced; they may be acoustically different from vowels in non-devoicing environments. If so, the overall process may be similar to the vowel reduction process in some other languages described in Chapter 2: high vowels are weakened between voiceless consonants, and they are finally reduced to being voiceless. \_

Chapter 7 examined the durations of voiced and voiceless morae, and found that voiceless morae were significantly shorter than voiced morae. However, voiceless morae were, at the same time, significantly longer than the corresponding consonants in the voiced morae. It was also found that, as far as plosives were concerned, durations of the closure periods in both voiced and voiceless morae were more or less the same but it was the duration of the release periods that differed significantly. The experiment showed that vowel devoicing is not ‘deletion’ as sometimes reported (Ohso, 1973, Kawakami, 1977, Beckman, 1982), but both the change of phonation from voiced to voiceless, and with a slight durational reduction.

The Japanese vowel weakening process seems to be different from the weakening processes in other languages because it may not involve obvious centralisation,

neutralisation or durational reduction. It is possible that the weakening process may work primarily on intensity reduction. In this chapter, I will measure the intensities of devoiced vowels in various environments, and analyse the environments for intensity reduction. Examining the factors affecting intensity reduction, I will consider the earlier proposal that vowel devoicing may be in fact a part of the weakening process. I will also discuss whether the above proposal can account for the remaining area of vowel devoicing, i.e. optionality and partialness of vowel devoicing in Japanese.

## 8.2 Experimental hypotheses

Chapters 5 and 6 investigated factors affecting vowel devoicing in Japanese. The experimental results in Chapter 5 found that the most important factors were (a) the type of preceding consonants, (b) devoicing condition, i.e. single or consecutive devoicing environments, and (c) position in an utterance (utterance final). Other proposed blocking factors, such as accentuation, presence of internal word boundary and position in an utterance other than utterance final, were not very effective in the single devoicing environment. In other words, high vowels were almost always devoiced as long as there were no devoiced vowels in adjacent morae. Chapter 6 examined the effect of speaking rate, and the results showed that there was no significant effect of speaking tempo on vowel devoicing rates in the single devoicing environment, but in the consecutive devoicing environment, devoicing occurred more often at faster tempi than slower tempi. It was also found that non-high vowel devoicing between voiceless consonants, and high vowel devoicing in non-devoicing environments, i.e. adjacent to or between voiced consonants, occurred more at faster tempi than slower tempi. However, frequencies of high vowel devoicing in the single devoicing environment were not affected by speaking tempo.

In the case of consecutive devoicing environments, however, not all vowels were devoiced. Devoicing in two consecutive morae sometimes occurred, but consecutive devoicing over two morae was uncommon at a normal speaking rate. Consecutive



high vowel devoicing in the devoicing environment over two morae occurred more at the fast tempo but rare at the normal and slow tempi. If it is natural for some of the devoiceable vowels in the consecutive devoicing environment to retain voicing, these devoiceable vowels may not necessarily be weakened.

The hypotheses of the present experiment are:

- (1) Voiced vowels in devoicing environments tend to be weaker than vowels in non-devoicing environments.
- (2) There is a difference in the degree of weakening, depending on the devoicing environment, since the probability of devoicing differs, depending on whether the vowels are in the single or consecutive devoicing environments.
- (3) The vowels are weakened more at faster tempi. This would be expected if devoicing is part of the vowel weakening process, resulting from vowel reduction similar to the phenomenon in other languages.

### **8.3 Experimental methods**

#### **8.3.1 Subjects**

Three Japanese speakers (one male and two female) who did not participate in other experiments in the previous chapters took part in the experiment. They were 32, 34 and 46 years of age, and speakers of Standard Japanese. They were not aware of the purpose of the experiment.

#### **8.3.2 Experimental settings**

An experiment was carried out to compare the intensities of vowels in voicing and devoicing environments. Vowel intensities were measured and compared under two experimental settings, namely (1) the devoicing condition, and (2) at 3 different speaking rates. Two environments were chosen for devoicing conditions, i.e. (a) the

single and (b) consecutive devoicing environments. Three speaking tempi were chosen for speaking rate, i.e. (i) fast, (ii) normal and (iii) slow.

### 8.3.3 Materials

#### 8.3.3.1 Test words

Six test words listed below were selected: (a) three with single devoicing sites and (b) three with single site and consecutive devoicing possibilities. Devoiceable vowels are underlined.

(a) /ta.i.sjo.k <u>u</u> .te`.a.te/	[taiçokwteate]	'retiring allowance'
/ka.mo.t <u>u</u> .se`.N.pa.k <u>u</u> /	[kamotsuɯsempakɯ]	'cargo boats'
/ta.ka.sa.ki.si`.mi.N/	[takasakiçimiŋ]	'the Takasaki citizen'
(b) /h <u>u</u> .k <u>u</u> .sjo.k <u>u</u> .ke`.N.sa/	[ɸɯkɯçokɯkeŋsa]	'dress inspection'
/sjo.k <u>u</u> .hi.se.tu.ja.k <u>u</u> /	[çokɯçisetsɯjakɯ] (no accent)	'a cut in food expenses'
/ha.i.sj <u>u</u> .t <u>u</u> .ki`.dju.N/	[haiçɯtsɯkidzɯŋ]	'exhaust limit'

N.B. /N/ is a moraic nasal, /sj/ is realised as [ç], /dj/ is realised as [dz], and /t/ followed by /u/ is realised as [ts].

The last /u/ underlined in (a) /kamotuseNpaku/ and (b) /sjokuhisetujaku/ are also in the devoicing environment (single sites), preceded by voiceless consonants and followed by /t/ of the carrier sentence /to (jomimasu)/, or followed by a pause when the subject inserted a pause between the test word and the carrier sentence. The test word (b) /hukusjokukeNsa/ has a single devoicing site (the underlined third /u/) as well as consecutive devoicing sites. The total number of devoiceable vowels in single devoicing sites in the test words was 6, and that in consecutive devoicing sites was also 6, yielding 324 devoiceable vowels ([6 vowels + 6 vowels] x 3 rates x 3 repetitions x 3 speakers = 324 devoiceable vowels).

The test words were embedded in a carrier sentence "kore wa \_\_\_\_\_ to yomimasu" ('This is pronounced \_\_\_\_\_').

### 8.3.3.2 Phonetic environments for the test words

The previous experiments in Chapters 5 - 7 showed that devoiced vowels were almost always devoiced in the single devoicing environment but not all devoiced vowels became voiceless in the consecutive devoicing environment. However, in this experiment, the intensities of *voiced* devoiced vowels in both single and consecutive devoicing environments were compared with those of non-devoiced vowels in the same words. Therefore, the following phonetic environments were chosen.

- (1) In the single devoicing environment, the devoiced vowels were always preceded by either voiceless plosives or affricates, because plosives and affricates sometimes triggered voicing from devoiced vowels, whereas devoiced vowels preceded by fricatives almost always became voiceless in both conditions (see Chapters 5 - 7).
- (2) In the consecutive devoicing environment, one of the devoiced vowels in the sequence was preceded by a fricative and the other was preceded by either a plosive or an affricate. The previous experiments showed that in the consecutive devoicing environment, devoiced vowels preceded by fricatives usually underwent a devoicing process, whereas devoiced vowels preceded by plosives and affricates tended to remain voiced. Therefore, one of the devoiced vowels in the consecutive devoicing environment might remain voiced in this experiment, which would enable us to compare intensities of vowels in devoicing and non-devoicing environments.

### 8.3.3.3 Speech rate setting

Each speaker pronounced each test word three times in random order at three different tempi: 'fast', 'normal' and 'slow'. The speakers adjusted their speaking tempi according to tones sounding in their headphones with a 3 second pause between sentences. The subjects were asked to time the first mora /ko/ of /kore/ of



the carrier sentence with the first beep, and the last mora /su/ of /yomimasu/ of the carrier sentence with the second beep; the first beep and second beep were 1.7 seconds apart for the 'fast' tempo, 2.5 seconds apart for the 'normal' tempo and 3.3 seconds apart for the 'slow' tempo.

In the previous experiment (Chapter 6), speaking rates were voluntarily controlled by the subjects. However, in this experiment, speaking rates were artificially manipulated by tones through headphones. This was done in order to control the actual speaking rate, and not just to allow the subjects to change the duration of pauses. An earlier study (Chapter 6) found a general tendency for devoicing to occur less at slower speaking rates. This was possibly because speakers articulated more precisely at slower rates, and so they did not skip the voicing of a vowel between voiceless sounds. However, another crucial reason was that speakers inserted a pause between phrases, sentences, and words. This created quite a number of devoiceable vowels occurring in utterance final position, which was one of the blocking factors of vowel devoicing. Adjusting speaking rates artificially by providing cues for the start and end timings for each sentence, minimised the possibility of speakers changing the duration of pauses, rather than adjusting actual speaking rates.

3 different tempi were set up in order (1) to examine intensity of vowels in devoicing sites at different tempi and (2) to trigger voicing of some vowels in devoicing sites. As an earlier study (Chapter 5 and Kondo, 1993) demonstrated, high vowels were almost always devoiced at normal speaking rate, if there was no devoiceable vowel in an adjacent mora. However, there was some effect of speaking tempo at the slow tempo (see Chapter 6). Therefore, devoiceable vowels may undergo a devoicing process at a normal speaking rate, but may remain voiced at a slower rate. Secondly, comfortable speaking rates vary from speaker to speaker. So the set "normal" speaking rate may not be the most comfortable rate for all subjects.

### 8.3.4 Intensity measurements

The average root mean square amplitude was calculated for each vowel by the ESPS Waves+ programme for a SUN workstation at a 16 KHz sampling rate. To obtain vowel intensities, the programme of 'power' under 'statistics' of ESPS Waves+ package was used. The 'power' function calculates the root mean square of the amplitude of vowels. With everything being equal, the longer a vowel, the greater its total intensity. Since the durations of vowels vary, the average intensity of each vowel, i.e. the total intensity of a vowel divided by its duration, was calculated. The beginning of vowels was defined as the onset of voicing and the end of vowels was defined as the point where periodic waveform ceased (for the criteria for deciding on segmental boundaries, see Chapter 5). The calculated average intensity of each vowel was then converted to decibels.

In order to compare the intensities of vowels in two different environments, the ratio between the intensity of each voiced devoiceable vowel, and the average intensity of non-devoiceable vowels in the same word for each speaker, excluding vowels in the first and last morae of each word, was calculated. The method of obtaining the intensity ratio was as follows:

- (i) The ratio between the average intensity of each voiced devoiceable vowel ( $I_1$ ) and the average intensity of non-devoiceable vowels ( $I_2$ ) was calculated.
- (ii) The logarithm of the obtained ratio in (iii) was calculated.
- (iii) The figure in (ii) was multiplied by 10, which gave a figure in decibels.

The formula is shown in (1) (Fry, 1979: 92):

$$(1) 10 \times \log \frac{I_1}{I_2} \text{ (dB)}$$

The intensities were compared only with data from the same subject, same tempo and within the same test word, in order to minimise the influence of the differences in recording and sampling levels. The devoiceable vowels were analysed under two conditions, namely in single or consecutive devoicing environments.

In actual utterances, intensity is influenced by various factors such as type of vowel, pitch, stress, and adjacent sounds. For instance, everything being equal, higher vowels, whose F1 and F2 are far apart, have lower intensities than lower vowels, whose F1 and F2 are close together. Similarly, a vowel at higher pitch or in stressed position has higher intensity than the same vowel at lower pitch or in unstressed position (for the detail, see Section 8.4.2). Therefore, for a pilot study, 14 nonsense words listed in (2i) and (2ii), for which phonetic environments were controlled in order to minimise the influence of factors affecting vowel intensity, were chosen. In other words, intensity of a voiced devoiceable vowel was compared with that of the non-devoiceable vowel of the same kind of the same pitch status in the same word.

(2) (i) Test words with the single devoicing environment for the pilot study

- |  |                                 |
|--|---------------------------------|
| a. /tak <i><b>i</b></i> s <u>i</u> de/ | b. /tak <u>u</u> s <u>i</u> de/ |
| c. /tap <i><b>i</b></i> s <u>i</u> de/ | d. /tap <u>u</u> s <u>i</u> de/ |
| e. /tat <i><b>i</b></i> k <u>i</u> de/ | f. /tat <u>u</u> k <u>i</u> de/ |

(2) (ii) Test words with the consecutive devoicing environment for the pilot study

- |  |  |
|--|--|
| a. /tak <i><b>i</b></i> s <i><b>i</b></i> k <u>i</u> ga/ | b. /tak <u>u</u> s <u>u</u> k <u>u</u> ga/ |
| c. /tap <i><b>i</b></i> s <i><b>i</b></i> p <u>i</u> ga/ | d. /tap <u>u</u> s <u>u</u> p <u>u</u> ga/ |
| e. /tat <i><b>i</b></i> s <i><b>i</b></i> t <u>i</u> ga/ | f. /tat <u>u</u> s <u>u</u> t <u>u</u> ga/ |
| g. /tah <i><b>i</b></i> s <i><b>i</b></i> h <u>i</u> ga/ | h. /tah <u>u</u> s <u>u</u> h <u>u</u> ga/ |

All the words in (2i) and (2ii) contained the high vowels /i/ and /u/ in both devoiceable (in bold italics) and non-devoiceable (underlined) positions. Since all the words were non-accented, both vowels were unaccented and high pitched. The experiment in Chapter 5 showed that in the single devoicing environment, devoiceable vowels were almost always devoiced, but less likely to be devoiced



when a vowel was preceded by a plosive or an affricate than when it was preceded by a fricative. Therefore, all devoiceable vowels were preceded by either a plosive or an affricate. In words with the consecutive devoicing environment listed in (2ii), one of the devoiceable vowels might have remained voiced, because in the consecutive devoicing environment, not all devoiceable vowels become voiceless. For (2ii) (a) ~ (f), the first high vowel was preceded by a plosive or an affricate and the second high vowel was preceded by a fricative, which was the more favoured environment for devoicing (see Chapter 5). Consequently, a devoiceable vowel preceded by a plosive may remain voiced, while the other devoiceable vowel preceded by a fricative may undergo the devoicing process. For (2ii) (g) ~ (j), both devoiceable vowels were preceded by fricatives. When two devoiceable vowels were preceded by fricatives, it was rare for both of the vowels to be devoiced (see Chapter 5). In each of the two words one of the devoiceable vowels was more likely to remain voiced than the other, which could have enabled comparison of the intensities of the voiced devoiceable and non-devoiceable high vowels. However, the subjects in the pilot study complained that they found it very difficult to pronounce the test words because they felt as if the words were tongue twisters. It was also very difficult find ordinary existing words with the same vowel occurring in two different environments, i.e. devoicing and non-devoicing, as well as high vowels occurring in the environment which triggers voicing, i.e. preceded by a plosive and followed by a fricative. Therefore, high vowel intensities, which tend to be lower, were compared with those of non-high vowel intensities, which tend to be higher.

In this experiment, the devoiceable high vowels in all the test words were high pitched, whereas the non-devoiceable vowels occurred in both high pitched and low pitched positions. The first vowel in each test word was low pitched when a subject inserted a pause between the carrier sentence, but it was high pitched when they did not insert a pause between them. All non-devoiceable vowels after the accented vowel in each word except in /sjokuhisetujaku/ which has no accent were low pitched. However, the intensities of the vowels in the first and last morae of each word were eliminated from data.

Although intensities of the vowels in the two environments were not compared under absolutely equal conditions, the data should show some tendency and still be a good indicator of vowel reduction. Three of the test words, /takasakisimiN/ with a single devoicing site, and /sjokuhisetujaku/ and /haisjutukidjuN/ with consecutive devoicing sites, contain the same high vowel in both devoicing and non-devoicing environments. Therefore, subsets of data from these 3 words were taken, and the intensity of a voiced high vowel in a devoicing environment was compared with the intensity of the same high vowel in a non-devoicing environment in the same word.

## 8.4 Results and discussion

### 8.4.1 Voicing rates

Table 8.1 shows the number of *voiced* vowels in the devoicing environment. As expected, the majority of devoiceable vowels were devoiced in the single devoicing environment for all subjects, but a high proportion of devoiceable vowels remained voiced in the consecutive devoicing environment. For subjects B and C, the majority of devoiceable vowels in the consecutive devoicing environment were in fact voiced. Subjects B and C showed high voicing rates even for words with only a single devoicing site, which was contradictory to the results in Chapter 5. This was possibly because the comfortable speaking rates of subjects B and C were faster than the set 'normal' tempo. Thus, they might have over-articulated in their pronunciations.

The total number of voiced devoiceable vowels pooled across the subjects was highest at the slow tempo and lowest at the fast tempo, but this was not always the case. Only subject C showed a neat pattern of the tempo effect for both conditions: more voicing at slower tempi and less voicing at faster tempi. Subject A voiced devoiceable vowels most at the fast tempo and did not voice at all at the normal tempo in the single devoicing environment. Subject B did not show any difference in voicing rates between fast and normal tempi in the single devoicing condition, and voiced most at the normal tempo in the consecutive devoicing condition.

**Table 8.1      The number of voiced vowels in single and consecutive devoicing environments**  
**(maximum number = 18 for both conditions at all tempi)**

subject	condition	Tempo			Total
		fast	normal	slow	
A	single	4	0	1	5/54 (9.26 %)
	consecutive	6	7	10	23/54 (42.59 %)
B	single	3	3	12	18/54 (33.33 %)
	consecutive	4	14	12	30/54 (55.56 %)
C	single	4	6	8	18/54 (33.33 %)
	consecutive	11	12	13	36/54 (66.67 %)
Total		32/108 (29.63 %)	42/108 (38.89 %)	56/108 (51.89 %)	130/324 (40.12 %)

The statistical analysis by ANOVA (2-way), taking the number of voiced vowels as the dependent variable and devoicing condition and tempo as factors, showed that the effect of the devoicing condition was significant [ $F(1, 12)=9.247$ ,  $p < .05$ ]. However, the effect of speaking tempo was not significant [ $F(2, 12)=1.935$ , n.s.]. There was no significant interaction between devoicing condition and tempo [ $F(2, 12)=.503$ , n.s.]. The statistical results showed, therefore, that when the actual speaking rate of the test words was controlled, the important factor for vowel (de)voicing was still the devoicing condition. Speaking tempo was not an important factor for (de)voicing frequencies. As stated earlier in this section, this may be the result of the artificially set speaking rates. In other words, subjects B and C may have found the set normal rate slow and the set slow rate extremely slow.

#### 8.4.2 Differences in vowel intensity

The intensities of voiced vowels in devoicing and non-devoicing environments were obtained. The average intensity of voiced vowels in non-devoicing environments was compared with the intensity of voiced vowels in devoicing environments in the



same word. Firstly, the intensities were analysed by subjects and speaking rates, and were compared using a T-test. The average intensities of vowels in non-devoicing environments, excluding word-initial and word-final morae, were calculated. Earlier experimental results in Chapters 5 and 6 suggested that there was a tendency for vowels in the initial and final morae in an utterance to be weakened. When a subject inserted a pause between "kore wa" of the carrier sentence and the test word, and/or between the test word and "to yomimasu", the vowels in the first and last morae of the test words became utterance-initial and utterance final, respectively. Therefore, all test word initial and test word final morae were excluded from the data. The underlined last /u/s in test words /kamotuseNpaku/ and /sjokuhisetujaku/ are in word final position and thus they should be excluded from the data. In fact, both /u/s in all examples by all subjects were completely voiceless, and so their intensities were not measured. Therefore, excluding these /u/s did not affect the result. Similarly, the underlined first /u/ in a test word /hukusjokukeNsa/ is in word initial position, and therefore it should be excluded from the data. However, it was also completely devoiced in all examples by all subjects, so excluding the initial /u/ in this test word did not affect the results either.

**Table 8.2    T-test results of intensity difference between vowels in devoicing environments and in non-devoicing environments (one-tailed)**

subject	tempo	average intensity of devoiceable vowels (dB)	average intensity of non-devoiceable vowels (dB)	df	p-value
A	fast	68.95	77.88	9	p < 0.001
	normal	72.37	75.60	6	p < 0.005
	slow	69.19	76.16	10	p < 0.005
B	fast	70.34	74.39	6	p < 0.05
	normal	75.51	78.32	16	p < 0.001
	slow	70.57	72.54	23	p < 0.001
C	fast	73.14	75.14	15	p < 0.05
	normal	73.34	74.81	15	n.s.
	slow	71.21	72.52	20	n.s.

Table 8.2 shows the average intensities of voiced devoiceable vowels and non-devoiceable vowels at three tempi by individual subjects. The intensities of voiced devoiceable vowels were significantly lower than those of ordinary vowels in the non-devoicing environment at all tempi, except for subject C at normal and slow tempi. The result was predicted because the earlier experimental results in Chapter 5 demonstrated that phonetic realisation of vowels in the devoicing environment varied from fully voiced to completely voiceless, and when a vowel happened to be voiced in the devoicing environment, it was often partially voiced, i.e. the duration of the vowel tended to be shorter and its intensity was lower (see Chapter 5 and Section 8.1 in this chapter).

The earlier experiment in Chapter 5 demonstrated that devoicing rates were largely dependent on the environment in which they occur; i.e. with single devoicing or consecutive devoicing environments. The experiment in Chapter 6 also demonstrated that the effect of speaking tempo on devoicing rates was significant in the single devoicing environment but not significant in the consecutive devoicing environment. Therefore, the intensities of vowels were compared by their environments using a T-test.

As shown in Table 8.3, when there was only one devoicing site in a word (the single devoicing condition), and the devoiceable vowels were voiced, the intensities of the devoiceable vowels were significantly lower than those of non-devoiceable vowels at all speaking tempi for all subjects. Subject A devoiced all devoiceable vowels at the normal tempo, so no comparison was made for the normal tempo. She voiced only one instance of a devoiceable vowel (/hukushokukeNsa/) at the slow rate, so the figures in brackets on the table show the raw intensities of the voiced devoiceable vowel and the average intensity of non-devoiceable vowels in the same word.

**Table 8.3 T-test results of intensity difference between vowels in single devoicing sites and non-devoicing environments (one-tailed)**

subject	tempo	average intensity of devoiceable vowels (dB)	average intensity of non-devoiceable vowels (dB)	df	p-value
A	fast	67.07	77.34	3	$p < 0.005$
	normal <sup>2</sup>	N/A	N/A	N/A	N/A
	slow <sup>3</sup>	(59.57)	(76.56)	N/A	N/A
B	fast	68.69	75.14	2	$p < 0.025$
	normal	72.63	77.64	2	$p < 0.05$
	slow	69.10	72.45	11	$p < 0.001$
C	fast	69.69	76.27	4	$p < 0.025$
	normal	71.10	74.85	5	$p < 0.05$
	slow	70.32	72.93	7	$p < 0.001$

With the consecutive devoicing environments presented in Table 8.4, however, when devoiceable vowels were voiced, their intensities were not necessarily lower. In the case of subject C at the fast tempo, the average intensity of devoiceable vowels was higher than those of non-devoiceable vowels. The statistical analysis by a T-test showed that the intensity of voiced devoiceable vowels and the intensity of non-devoiceable vowels were significantly different at all tempi for subject A and at the normal tempo for subject B, but the differences at the fast and slow tempi for subject B and at all tempi for subject C were not found to be significant.

---

<sup>2</sup> Subject A devoiced all devoiceable vowels.  
<sup>3</sup> There was only one instance of a voiced vowel in the devoicing site.



**Table 8.4 T-test results of intensity difference between vowels in consecutive devoicing environments and in non-devoicing environments (one-tailed)**

subject	tempo	average intensity of devoiceable vowels (dB)	average intensity of non-devoiceable vowels (dB)	df	p-value
A	fast	70.20	78.25	5	$p < 0.025$
	normal	72.37	75.60	6	$p < 0.005$
	slow	70.15	76.12	9	$p < 0.005$
B	fast	71.57	73.83	3	n.s.
	normal	76.12	78.47	13	$p < 0.001$
	slow	72.03	72.63	11	n.s.
C	fast	74.71	74.63	10	n.s.
	normal	74.46	74.79	11	n.s.
	slow	71.76	72.26	12	n.s.

In languages such as English, the intensity of a vowel in a stressed syllable is greater than the intensity of the same vowel in an unstressed syllable, suggesting that the increased vocal effort of stressed syllables produces a simple intensity difference (Pickett, 1980: 181). Languages such as Japanese do not have an obvious stress, but accentuation involves pitch alternation in ordinary utterances, and intensity is also influenced by the change of pitch; if there are two sounds of equal amplitude, one with a higher frequency than the other, then more energy will be needed for the higher frequency since the to-and-fro motion of the vocal folds is taking place more often, resulting in higher intensity (Fry, 1979: 90).

As discussed earlier in Section 8.3.4, different vowels have their own intrinsic intensities even when spoken with equal effort. Vowels made with a wider vocal tract have a higher intensity level than the close vowels. For the same degree of opening, vowels with closer F1 and F2 have a higher intensity than vowels with F1 and F2 far apart; i.e. back vowels are a little more intense than front vowels, and for

English the difference between the highest and the lowest intensity vowel sound is on average 7 dB (Fry, 1979: 117, Pickett, 1980: 181). Fry (1979: 126-127) listed the intensity of English sounds with reference to a consonant [θ], which is naturally the weakest sound in English. The intensities of English vowels relative to [θ] in decibels are:

**Table 8.5 Relative intensities of English vowels with reference to [θ] in decibels**

(from Fry, 1979: 127)

[i:] 22	[e] 23	[u] 24	[ə:] 25	[ʌ] 26	[o] 28	[o:] 29
[u:] 22		[a] 24		[a:] 26		
[i] 22						

As far as the positions of F1 and F2 of Japanese vowels are concerned, F1 and F2 of the vowels [i], [e] and [ɯ] are relatively far apart while [a] and [o] have relatively close F1 and F2 (the formant frequencies are based on Delattre et al., 1952, cited in Borden and Harris, 1980: 178). In other words, the intensities of [i], [e] and [ɯ] are generally less than those of [a] and [o]. In this experiment, all devoiceable vowels were either [i] or [ɯ] which are inherently weak in intensity. This may have lowered the average intensity ratios of devoiceable vowels against non-devoiceable vowels which are inherently greater in intensity.

Moreover, as pitch changed within a word, some vowels in the test words were on a higher pitch than others, which would have affected their intensity ratios. When the test words were pronounced in isolation, the first mora of the all test words was on a low pitch and the last two morae of all words were also low pitched, except for /sjokuhisetujaku/ where only the first mora is low pitched as the word is not accented. Although the vowels in the first and last morae of each word were eliminated from the statistics, the above intensity ratios were not a true comparison under strictly equal conditions.

Furthermore, adjacent consonants also influence the intensity of vowels. Each consonant has its inherent intensity, and consonants are weaker in intensity than vowels. Table 8.6 lists the intensities of English consonants with reference to [θ]. Generally speaking, semi-vowels and liquids have higher intensity, followed by nasals. Fricatives and plosives tend to have lower intensity except for [ʃ] which has a markedly wide noise-band from fairly low to very high frequencies and has relatively high intensity among the consonants. Although Japanese consonants differ from English consonants, intensities of Japanese consonants should show similar tendencies to English consonants.

**Table 8.6    Relative intensities of English consonants with reference to [θ] in decibels**

(from Fry, 1979: 127)

[w]	21	[l]	20	[m]	17	[dʒ]	13	[s]	12	[k]	11	[b]	8	[f]	7
[r]	20	[ʃ]	19	[tʃ]	16	[ʒ]	13	[t]	11	[v]	10	[d]	8		
[j]	20	[ŋ]	18	[n]	15	[z]	12	[g]	11	[ð]	10	[p]	7		

As for the influence of adjacent consonants on vowel intensity, nasalisation for instance has the effect of broadening the frequency bandwidth and reducing the intensity of the oral formants relative to those seen near the stop consonant closures (Pickett, 1980: 126). Some of the non-devoiceable vowels in the test words were adjacent to or surrounded by nasals and inevitably were nasalised. The nasalisation also might have brought the average intensity of the non-devoiceable vowels down. On the other hand, voiceable vowels were always surrounded by voiceless obstruents which have generally weaker intensity than nasals, liquids and semi-vowels. Non-devoiceable vowels were often adjacent to nasals, liquids and semi-vowels as well as voiced obstruents. If the intensity of adjacent consonants has some effect, either raising or lowering the intensity of the vowels, voiceless obstruents adjacent to the voiceable vowels may have lowered vowel intensities.



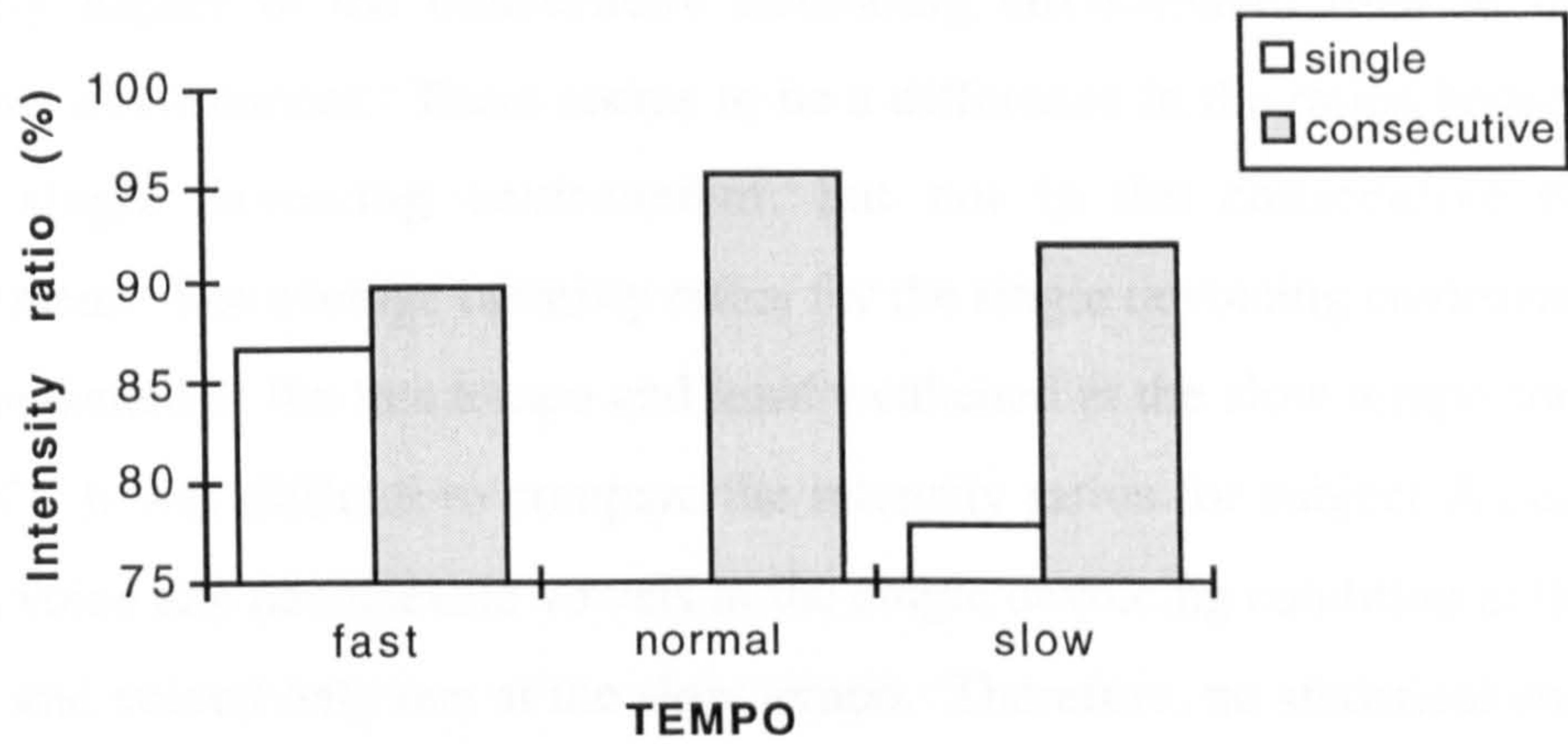
### 8.4.3 The effect of speaking rate on vowel intensity

Earlier results (Chapter 6 and Section 8.4.1 in this chapter) showed that speaking tempo did not necessarily influence vowel devoicing rates. The experiment in Chapter 6 showed that when speaking tempo was controlled by speakers, tempo had a significant effect on devoicing rates in the single devoicing environment but was not effective in the consecutive devoicing environment. However, there might have been factors other than tempo, such as the presence of (internal-) word boundaries and subsequently an increased number of pauses, affecting devoicing rates indirectly (see Chapter 6, Section 6.4.7). The experimental result in this chapter showed that when speaking tempo was controlled artificially, although there was a tendency for devoiceable vowels to be voiced more at slower tempi, the effect of tempo on vowel voicing rates was not statistically significant (see Section 8.4.1).

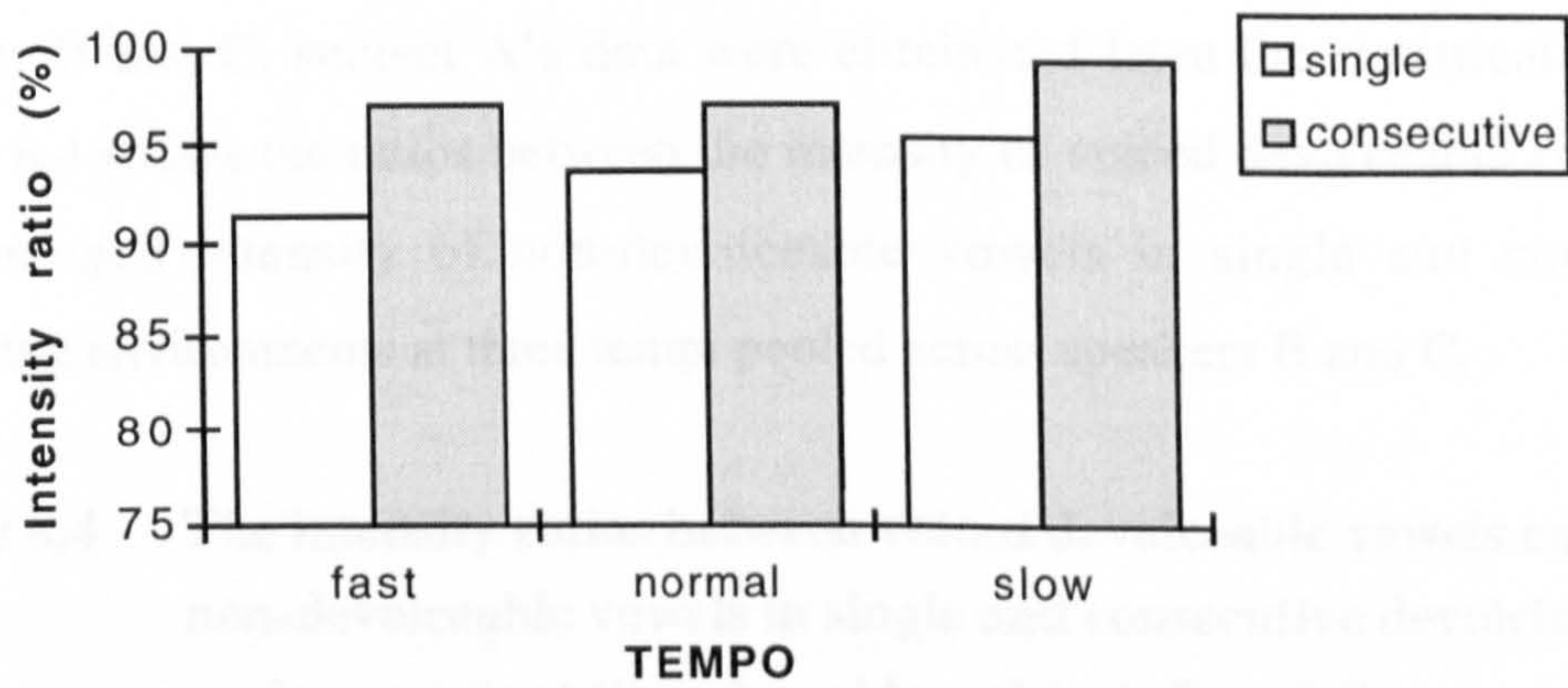
Vowel (de)voicing rates were not always affected by speaking tempo, especially in the consecutive devoicing environment. That implies that the vowel devoicing process in Japanese may be different from vowel reduction processes reported in other languages (see Chapter 2). On the other hand, as mentioned earlier in Chapter 5 and Section 8.1 in this chapter, phonetic realisations of devoiceable vowels differ greatly, from completely voiceless to fully voiced. That means that examining (de)voicing rates only cannot reflect the true features of the tempo effect in the vowel devoicing process. Therefore, in this section, the intensities of voiced devoiceable vowels and of vowels in non-devoicing environments were examined. In order to compare the intensities of vowels in two different environments, the ratio between the intensity of each voiced devoiceable vowel and the averaged intensity of non-devoiceable vowels excluding vowels in the first and last morae was calculated. The comparison was made only within the same word, for the same speaker, for all tempi, in order to minimise the influence of the difference in recording and sampling levels. The devoiceable vowels were analysed under two conditions, namely in single or consecutive devoicing environments. The results are shown in Figures 8.3a, 8.3b and 8.3c.

**Figure 8.3** The intensity ratios between vowels in devoicing environments and vowels in non-devoicing environments in single and consecutive devoicing environments at three tempi by individual subjects

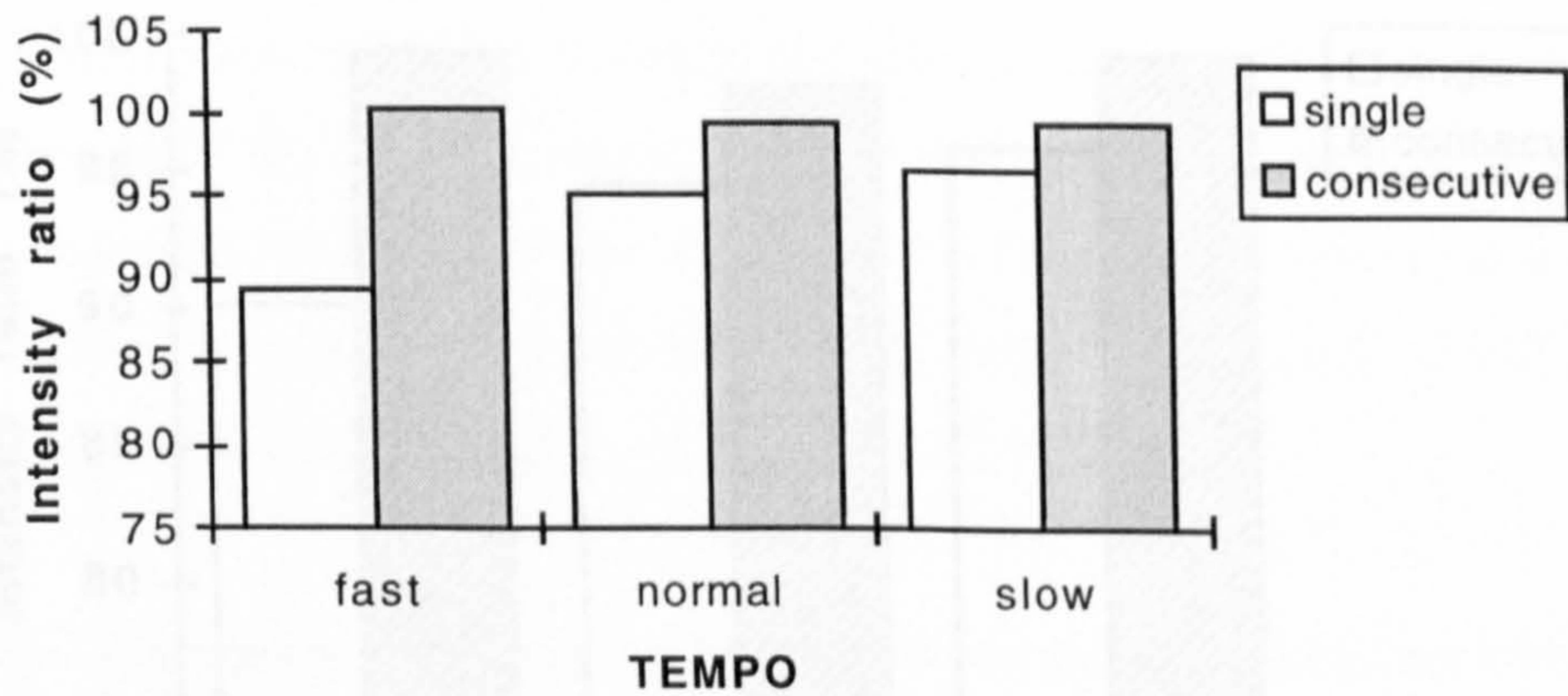
(a) Subject A<sup>4</sup>



(b) Subject B



(c) Subject C

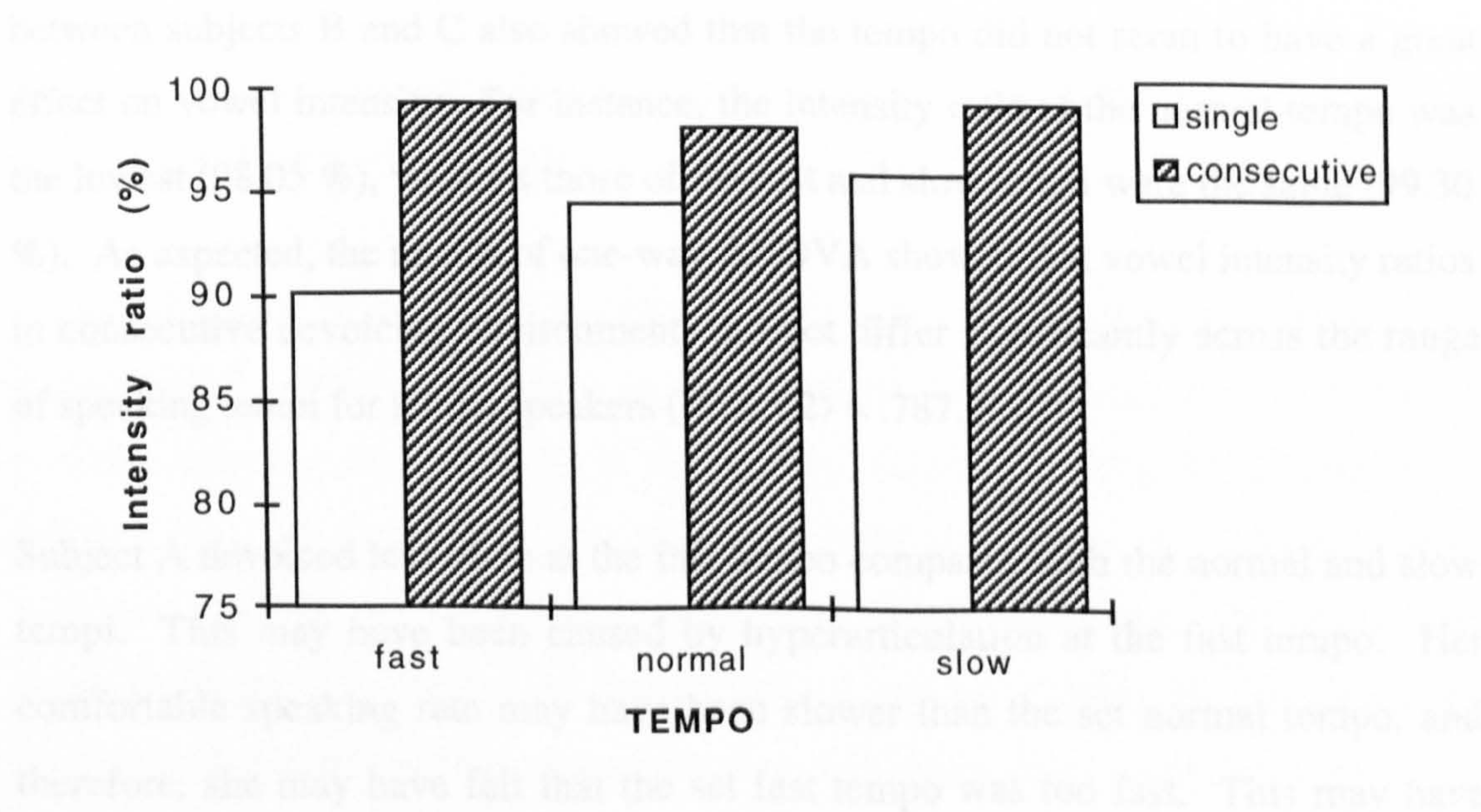


<sup>4</sup> Subject A did not voice devoiceable vowels at the normal tempo. So no comparison was made. There was only one sample of voiced high vowels in the devoicing environments at the slow tempo.



Although the statistical results were not significant, there was a general tendency for devoiceable vowels to remain voiced at slower tempi as shown in Table 8.1. The intensity ratios obtained showed a similar pattern. The intensity ratios were generally higher in the consecutive devoicing environment than in the single devoicing environment. There seems to be a difference in the ratios between tempi in the single devoicing environment, but not in the consecutive devoicing environment. The average intensity ratios for the single devoicing environment were most weakened at the fast tempo and least weakened at the slow tempo for subjects B and C. It was difficult to compare the intensity ratios for subject A because she did not voice any devoiceable vowels in the single devoicing condition at the normal tempo, and voiced only one at the slow tempo. Therefore, no statistical comparison was made for the ratios at the normal and slow tempi in the single devoicing environment for subject A. As subject A showed different voicing patterns from subjects B and C, subject A’s data were eliminated from the statistical analyses. Figure 8.4 shows the ratios between the intensity of voiced devoiceable vowels and the averaged intensity of non-devoiceable vowels in single and consecutive devoicing environments at three tempi pooled across speakers B and C.

**Figure 8.4     The intensity ratios between voiced devoiceable vowels and non-devoiceable vowels in single and consecutive devoicing environments at three tempi by subjects B and C**





A two-way analysis of variance taking intensity ratio as the dependent variable and devoicing conditions (single or consecutive) and speaking tempo as factors showed that there were significant effects from both factors: environments [ $F(1, 96) = 41.807, p < .0001$ ], and tempo [ $F(2, 96) = 3.880, p < .025$ ]. There was a significant interaction between environments and tempo [ $F(2, 96) = 4.252, p < .025$ ]. In other words, the voiced devoiceable vowels were weakened significantly in the single devoicing environment, and weakened more at faster tempi.

The effect of tempo was not shown on the voicing rates in Section 8.4.1, but it was clearly reflected in the intensity of voiced devoiceable vowels. In the single devoicing environment, devoiceable vowels were voiced, but they were acoustically different from vowels in the non-devoicing environment. Voiced devoiceable vowels were weakened significantly at faster tempi. The results suggest that vowel devoicing in Japanese may be part of vowel weakening process which is sensitive to speaking tempo as predicted, although it was not always shown clearly in the vowel devoicing rates.

In the consecutive devoicing environments, however, the intensities of the voiced devoiceable vowels were not greatly weakened. Some of their intensities were, in fact, higher than those of the non-devoiceable vowels. The averaged intensity ratio between subjects B and C also showed that the tempo did not seem to have a great effect on vowel intensity. For instance, the intensity ratio at the normal tempo was the lowest (98.05 %), whereas those of the fast and slow tempi were the same (99.30 %). As expected, the results of one-way ANOVA showed that vowel intensity ratios in consecutive devoicing environments did not differ significantly across the range of speaking tempi for all the speakers ( $[F(2, 62) = .787, \text{n.s.}]$ ).

Subject A devoiced less often at the fast tempo compared with the normal and slow tempi. This may have been caused by hyperarticulation at the fast tempo. Her comfortable speaking rate may have been slower than the set normal tempo, and therefore, she may have felt that the set fast tempo was too fast. This may have

made her articulate better resulting in less vowel devoicing in the fast tempo utterances than in the normal tempo utterances.

The vowel devoicing process in Japanese seems to be different from similar phenomena in other languages such as English, German, Greek and Hebrew, where vowels may be devoiced during the process of vowel reduction in fast speech (Gimson, 1980, Kohler, 1990, Dauer, 1980a & 1980b, Semiloff-Zelasko, 1973, etc. and Chapter 2). Vowel devoicing in Japanese occurred at a slow speaking rate as well as at a fast speaking rate as demonstrated in Chapter 6. However, the effect of devoicing conditions, namely single or consecutive devoicing environments, was very crucial. Vowel devoicing is not a compulsory process but seemed to be a favoured process in single devoicing environments. Devoiceable vowels in the single devoicing environment could be voiced, but their intensities were smaller than those of non-devoiceable vowels. Similar to the results of devoicing rates, the effect of speaking tempo on the intensity ratio differed depending on the devoicing environments. The degree of intensity weakening was greater at faster tempi than at slower tempi in the single devoicing environment, but there was no effect of speaking tempo on the intensity ratios in the consecutive devoicing environment. The results suggested that different vowel weakening processes seem to operate in single and consecutive devoicing environments.

#### **8.4.4 Intensity ratios of voiced vowels of the same kind in devoicing and non-devoicing environments**

As stated earlier in Section 8.3, vowels have their own inherent intensities. Also phonological environments such as neighbouring sounds, pitch and accentuation strongly influence the intensities of the vowels. Due to the difficulty in finding test words ideal for comparing intensities in both devoicing and non-devoicing environments, the type of vowel tested was not always the same. Although there were differences between the intensities of voiced vowels in the devoicing environment and non-devoicing environment, this might simply have been due to the

different types of vowels in the two environments. Under equal conditions, high vowels have intrinsically lower intensities than non-high vowels, and all vowels in the devoicing environment are high vowels. Therefore, intensity ratios of the subset of the test words were compared.

Three words among the six test words have the same type of vowels in the two devoicing environments: namely, /ta.ka.sa.ki.si'.mi.N/ 'the Takasaki citizen', /sjo.ku.hi.se.tu.ja.ku/ 'a cut in food expenses' and /ha.i.sju.tu.ki'.dju.N/ 'exhaust limit'. In all test words, the devoiceable vowels are underlined. /ta.ka.sa.ki.si'.mi.N/ has a devoiceable /i/ in a single devoicing environment and other /i/s in non-devoicing environments. /sjo.ku.hi.se.tu.ja.ku/ has two /u/s in devoicing environments, the first devoiceable /u/ in a consecutive devoicing environment and the second devoiceable /u/ in a single devoicing environment, as well as another /u/ in /tu/ in a non-devoicing environment. However, all three subjects devoiced all instances of the word final /u/ in the single devoicing environment. Thus the comparison of the intensities will be made only between the first devoiceable /u/ in a consecutive environment and the other /u/ in a non-devoicing environment. /ha.i.sju.tu.ki'.dju.N/ has two devoiceable /u/ in consecutive devoicing environments and another /u/ in /dju/ in a non-devoicing environment. Accentuation and neighbouring sounds still differed and the number of samples was small, especially in the single devoicing environment, because devoiceable vowels in the single environment were usually devoiced and the number of voiced samples of devoiceable vowels in the single environment were limited as shown in Table 8.5. However, intensity measurements of the subset may provide more realistic results.

First of all, the intensity ratios of vowels of the same kind in the subset between in the devoicing and non-devoicing environments in single and consecutive devoicing environments were compared at three tempi by individual subjects. As shown in Figure 8.7, the results of the intensity ratios for the subset showed similar patterns as the intensity ratios for the whole set of test words shown in Figure 8.3. The intensities of voiced devoiceable vowels in the single devoicing environment were generally weaker than those in the consecutive devoicing environment. For the



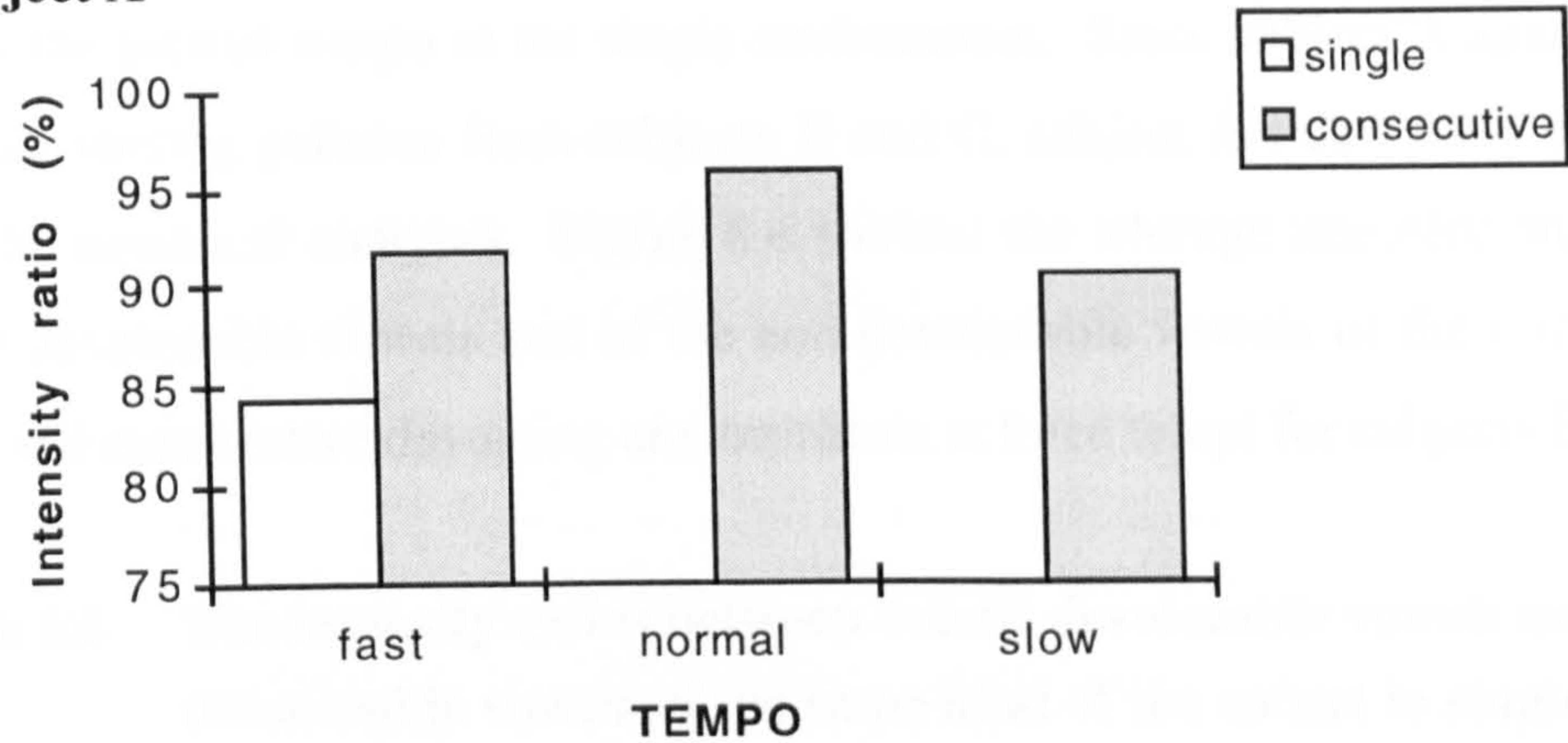
subset, Subject A did not voice voiced devoiceable vowels at the normal and slow tempi in the single devoicing environment, and Subject B devoiced only one example of devoiceable vowel at the fast tempo in the single devoicing environment.

**Table 8.7    The number of voiced devoiceable vowels of the same kind in the subset in the single and consecutive devoicing environmets**

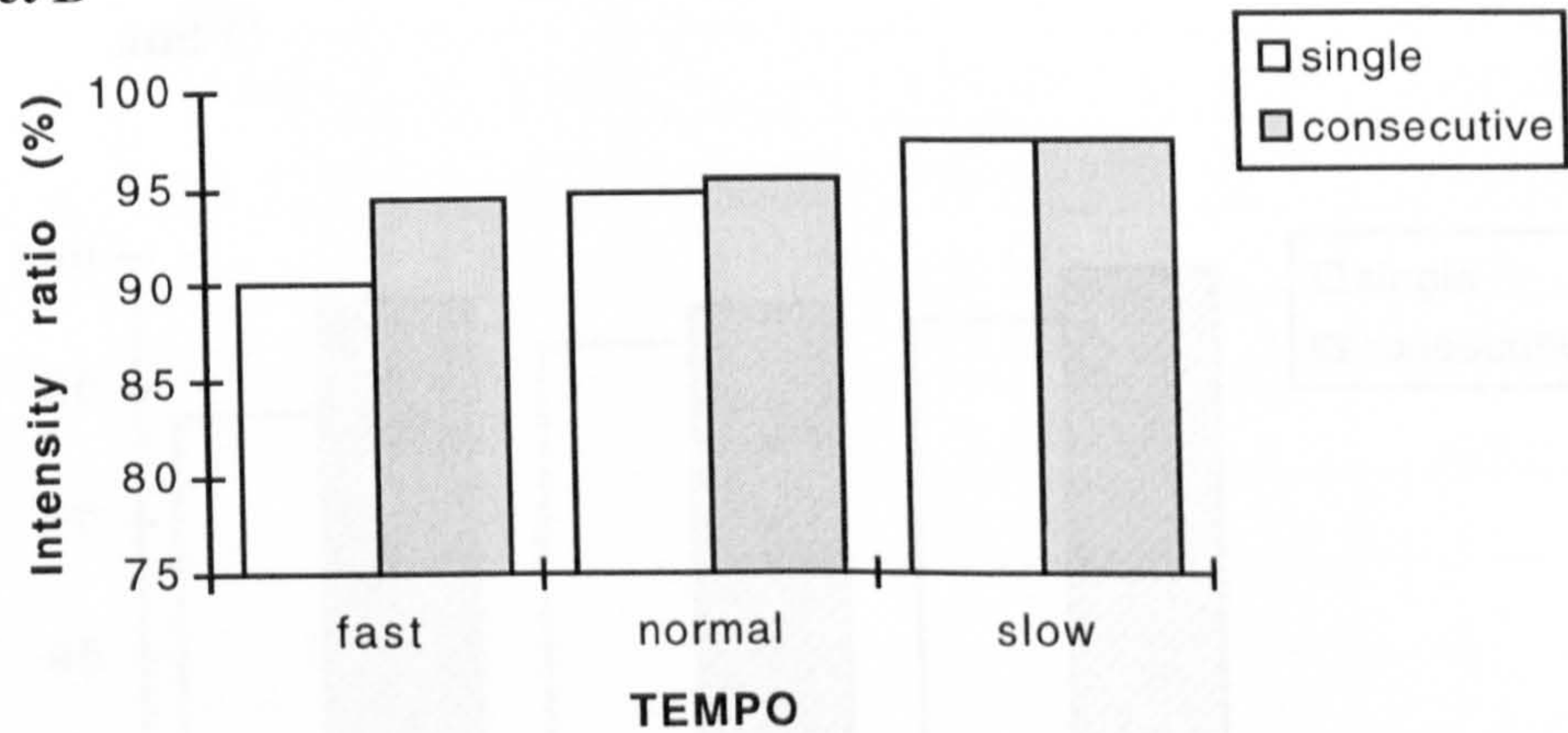
	tempo	the number of voiced devoiceable vowels in the single devoicing condition of the subset	the number of voiced devoiceable vowels in the consecutive devoicing condition of the subset
A	fast	2	4
	normal	0	3
	slow	0	6
B	fast	1	4
	normal	2	6
	slow	3	6
C	fast	3	5
	normal	3	6
	slow	3	6

**Figure 8.5** The intensity ratios between vowels in the subset in devoicing environments and vowels in non-devoicing environments in single and consecutive devoicing environments at three tempi by individual subjects

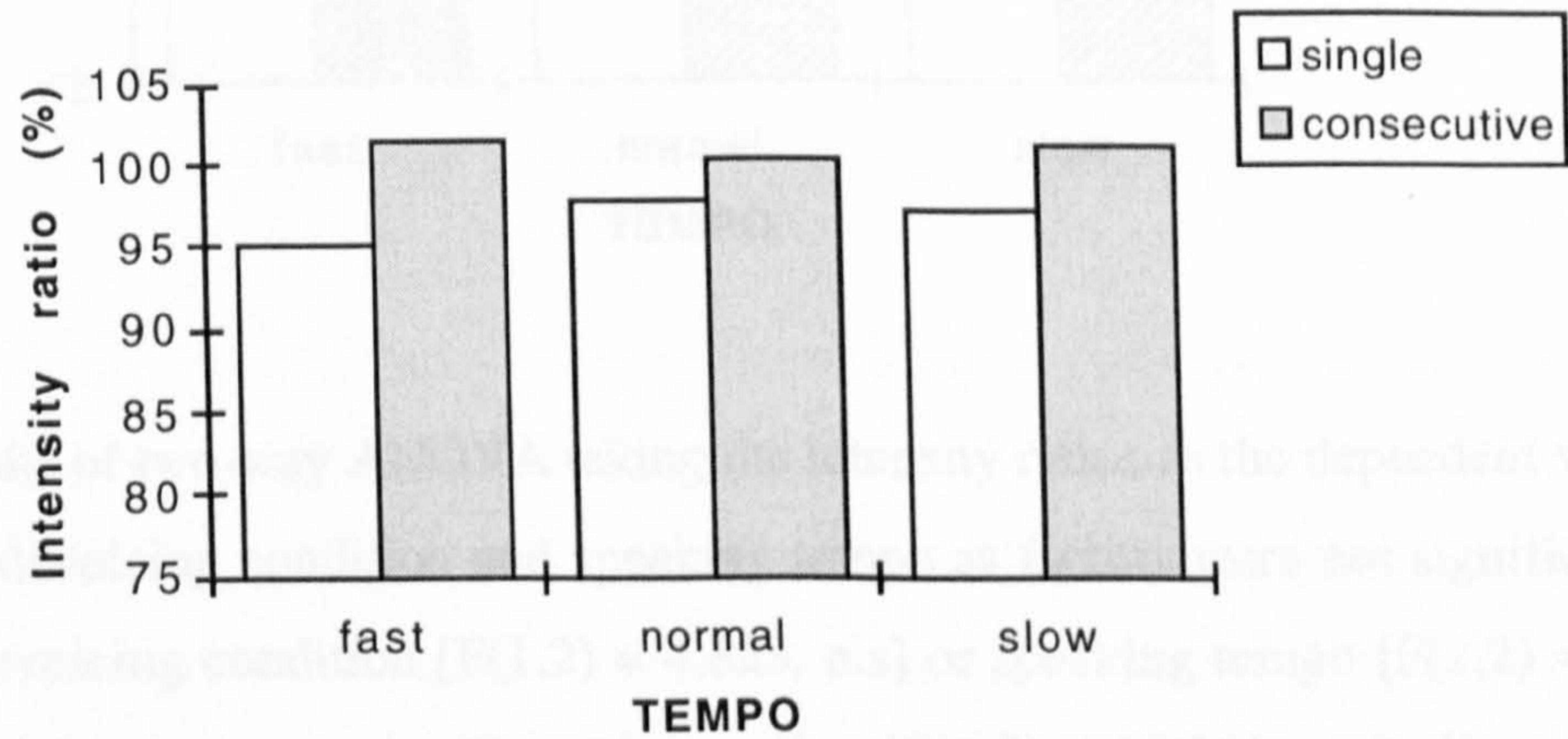
**(a) Subject A<sup>5</sup>**



**(b) Subject B<sup>6</sup>**



**(c) Subject C**



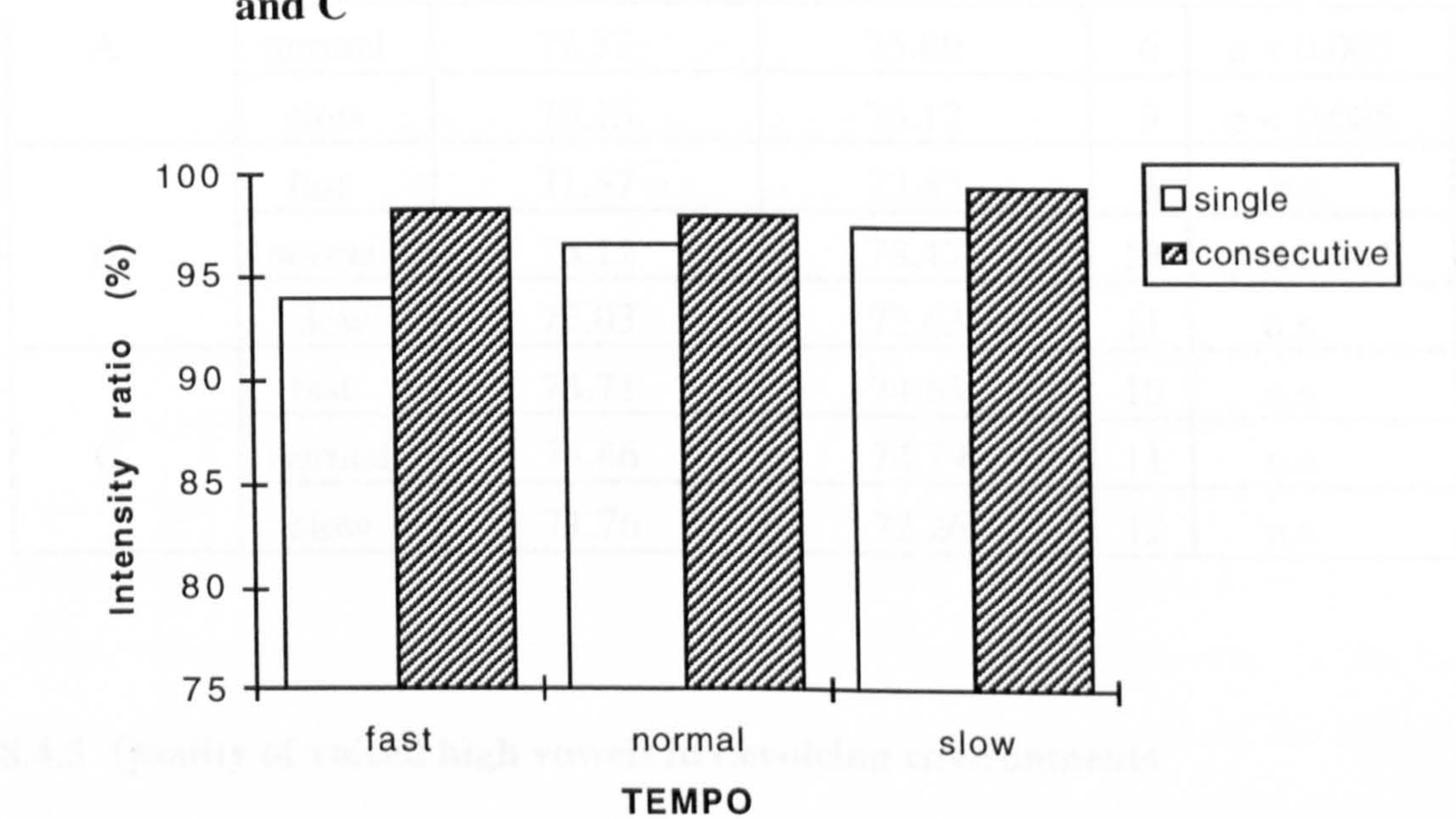
<sup>5</sup> Subject A did not voice devoiceable vowels at the normal and slow tempi in the single devoicing environment. Therefore no comparisons were made for these two environments. There was only one sample of voiced high vowels in the devoicing environments at the slow tempo.

<sup>6</sup> The intensity ratio at the fast tempo in the single devoicing environment was based on one data.



For subject B, the average intensity ratios were weakened most at the fast tempo and weakened least at the slow tempo, for both single and consecutive environments. For subject C, the average intensity ratios were weakened most at the fast tempo, but least at the normal tempo in the single environment. Since subject A again showed different voicing patterns from subjects B and C, subject A's data were eliminated from the statistical analyses. Figure 8.6 shows the average intensity ratios of the voiced devoiceable vowels and of the non-devoiceable vowels of the same kind in single and consecutive devoicing environments at three tempi for subjects B and C.

**Figure 8.6** The intensity ratios between voiced devoiceable vowels and non-devoiceable vowels of the same kind of the subset in single and consecutive devoicing environments at three tempi by subjects B and C



The results of two-way ANOVA taking the intensity ratios as the dependent variable and the devoicing condition and speaking tempo as factors were not significant for

either devoicing condition [ $F(1,2) = 4.823$ , n.s] or speaking tempo [ $F(2,2) = 2.056$ , n.s.], and there was no significant interaction [ $F(2,2) = 10.261$ , n.s]. However, the patterns of the intensity ratios for the subset were very similar to those of the whole set shown in Figure 8.4. The following patterns were also noted for the subset: (a) more intensity weakening in the single devoicing environment than in the



consecutive environment at all tempi, (b) most intensity weakening at the fast tempo and least intensity weakening at the slow tempo in the single devoicing environment, and (c) there seemed to be no tempo effect on intensity in the consecutive devoicing environment. The lack of significant results may have been due to the low number of replicates in the subset.

**Table 8.8 T-test results of intensity difference between vowels in consecutive devoicing environments and in non-devoicing environments (one-tailed)**

subject	tempo	average intensity of devoiceable vowels (dB)	average intensity of non-devoiceable vowels (dB)	df	p-value
A	fast	70.20	78.25	5	$p < 0.025$
	normal	72.37	75.60	6	$p < 0.005$
	slow	70.15	76.12	9	$p < 0.005$
B	fast	71.57	73.83	3	n.s.
	normal	76.12	78.47	13	$p < 0.001$
	slow	72.03	72.63	11	n.s.
C	fast	74.71	74.63	10	n.s.
	normal	74.46	74.79	11	n.s.
	slow	71.76	72.26	12	n.s.

#### 8.4.5 Quality of voiced high vowels in devoicing environments

The results of vowel intensity measurements showed that high vowels may remain voiced between voiceless consonants, but they tend to be reduced in intensity. Earlier results in Chapters 5 and 7 also indicated that voiced devoiceable vowels tended to be shorter in duration. These results suggest that vowel devoicing may be part of the vowel weakening process involving durational reduction and intensity weakening. Vowels may be reduced both in duration and intensity, but does the process also involve a change of vowel quality such as centralisation and

neutralisation similar to the processes reported in other languages? In this section, the quality of vowels in various environments will be examined relation to the reduction of intensity and duration.

The first, second and third formants of all examples of the voiced high vowels /i/ and /u/ in both devoicing and non-devoicing environments were measured. The first and second formant frequencies (F1 and F2 respectively) were plotted on linear scale graphs according to the devoicing environments in which they occurred. This was because the different tongue shapes and positions are reflected in the frequency positions of particularly the first and second formants, and listeners' auditory judgement in identifying vowels is also based on the relationship between F1 and F2 (Borden and Harris, 1984: 177, Fry, 1979: 78, Pickett, 1980: 170, etc.).

Vowels in both devoicing and non-devoicing conditions did not appear in precisely the same environment in the test words, which made it difficult to compare formant frequencies of the vowels because formant frequencies are influenced greatly by neighbouring sounds. There were some problems in measuring formant frequencies: (a) all high vowels in devoicing environments inevitably occurred either between voiceless obstruents or between a voiceless obstruent and a pause, whereas the adjacent sounds for non-devoiceable high vowels varied enormously; vowels, voiced obstruents, voiceless obstruents, nasals, semi-vowels, and a liquid, (b) when a high vowel was adjacent to another vowel or a semi-vowel, its influence on the high vowel was very strong; there was a very long transitional period of the formants from one vowel to another, and the formants did not achieve the target frequencies virtually until the end of the high vowel in question, (c) some of the /i/ vowels were heavily nasalised as they were either adjacent to or surrounded by nasals, which made measuring F1 and F2 complicated, because nasalisation results in an anti-resonance in the region of F1 and F2 (Pickett, 1980: 153), and (d) several vowels in devoicing environments were extremely short and weak, with the periodic waveform repeated only a few times.

In order to minimise the influence of neighbouring sounds on the vowel formants, the following criteria were used for the formant measurements:

- (a) Formants were measured in the middle of a steady state, regardless of the length of the steady period: this steady state was sometimes very short at the beginning or the end of vowels.
- (b) When a vowel had a fairly steady formant transition throughout its duration, its formants were measured at the highest point of the amplitude because this corresponds to the perceptual judgement of vowels.
- (c) Even if a vowel was very short and weak, the formants were measured in the middle of the vowel portion.

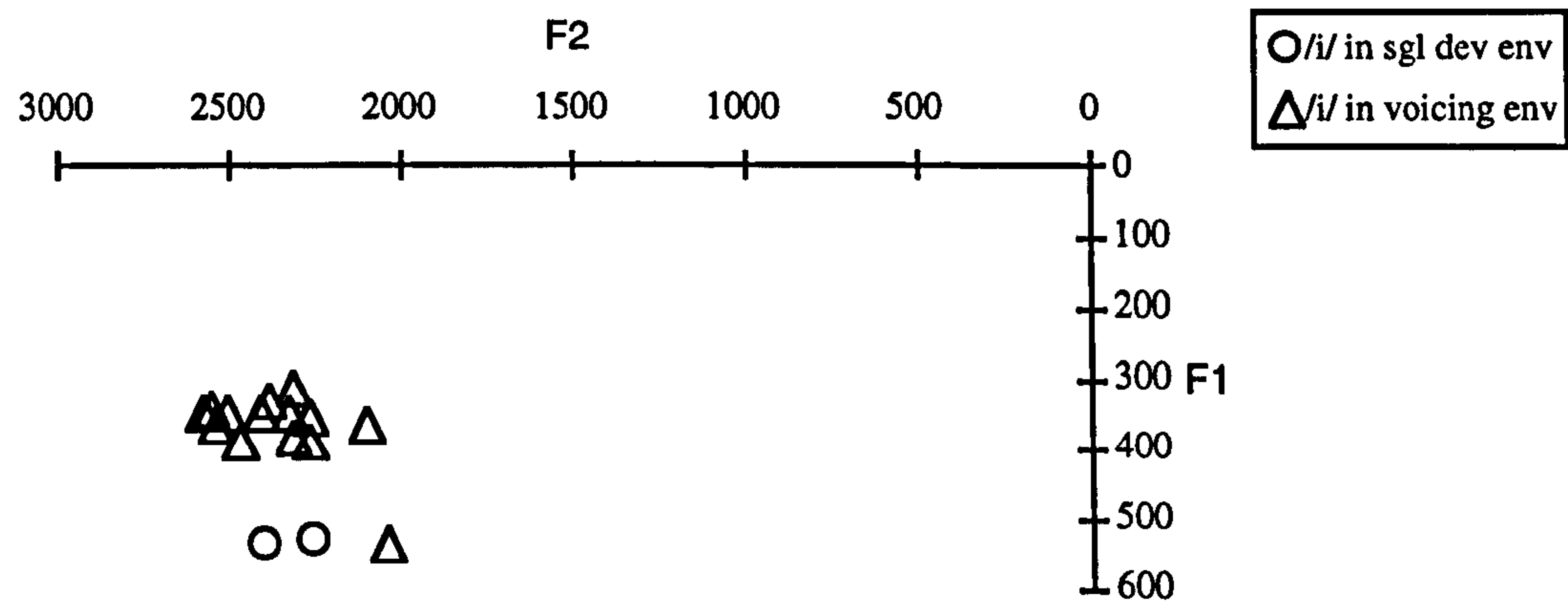
F1 and F2 are plotted in Figures 8.7 (a~c) ~ 8.12 (a~c).



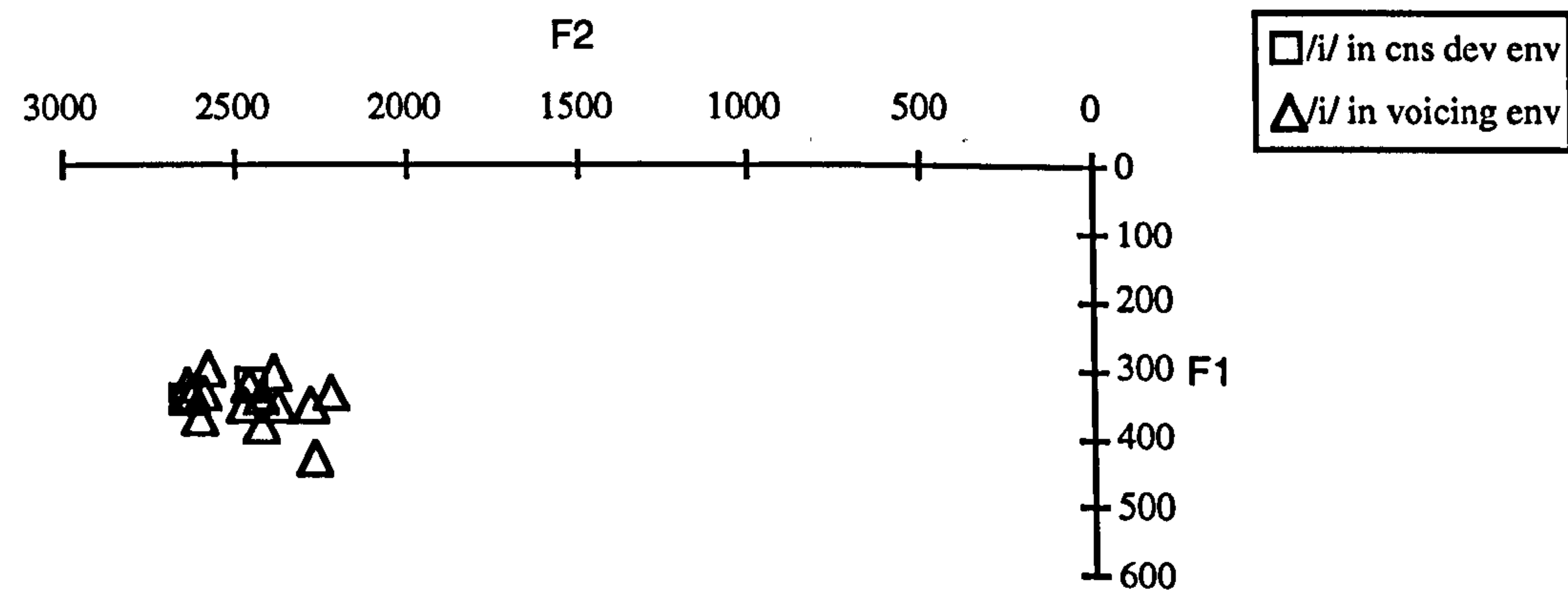
Figures 8.7a ~ 8.7c

F1 and F2 of individual tokens of /i/ in the consecutive devoicing, single devoicing and non-devoicing environments at different tempi by subject A (female)

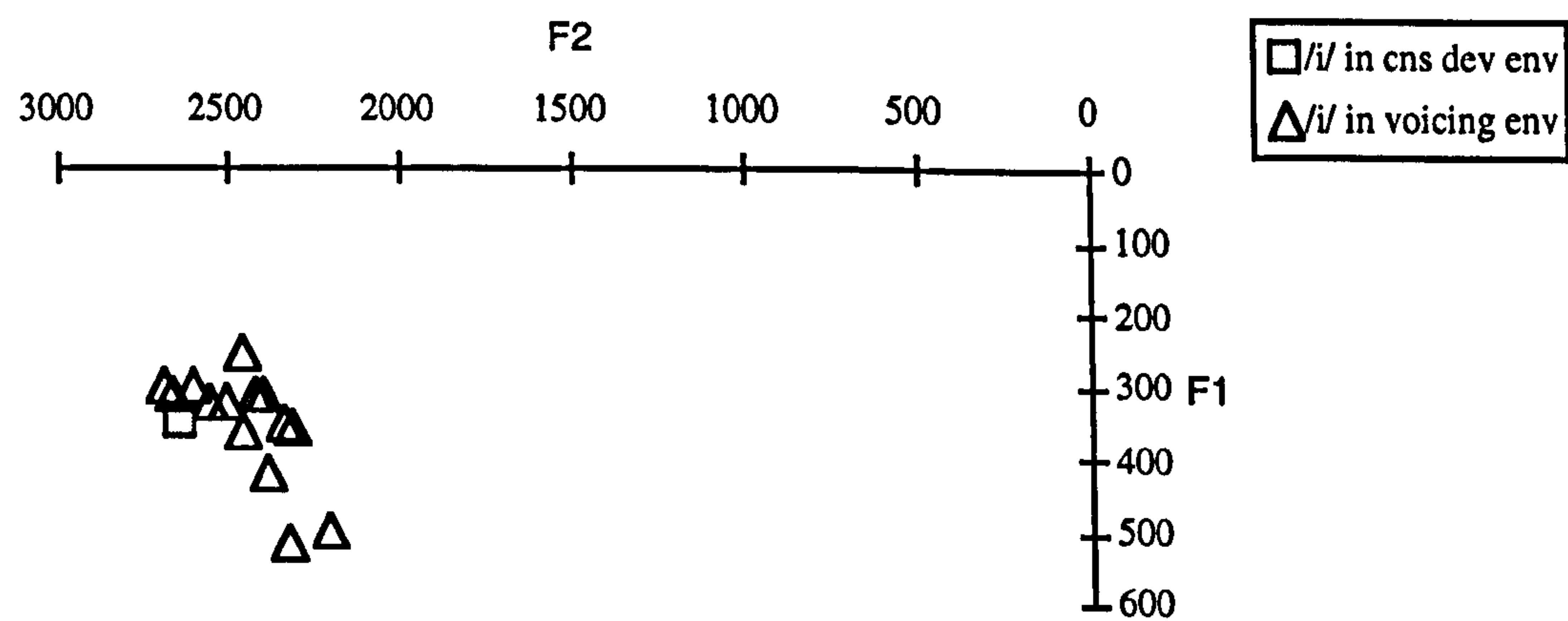
(a) A: Fast tempo /i/



(b) A: Normal tempo /i/



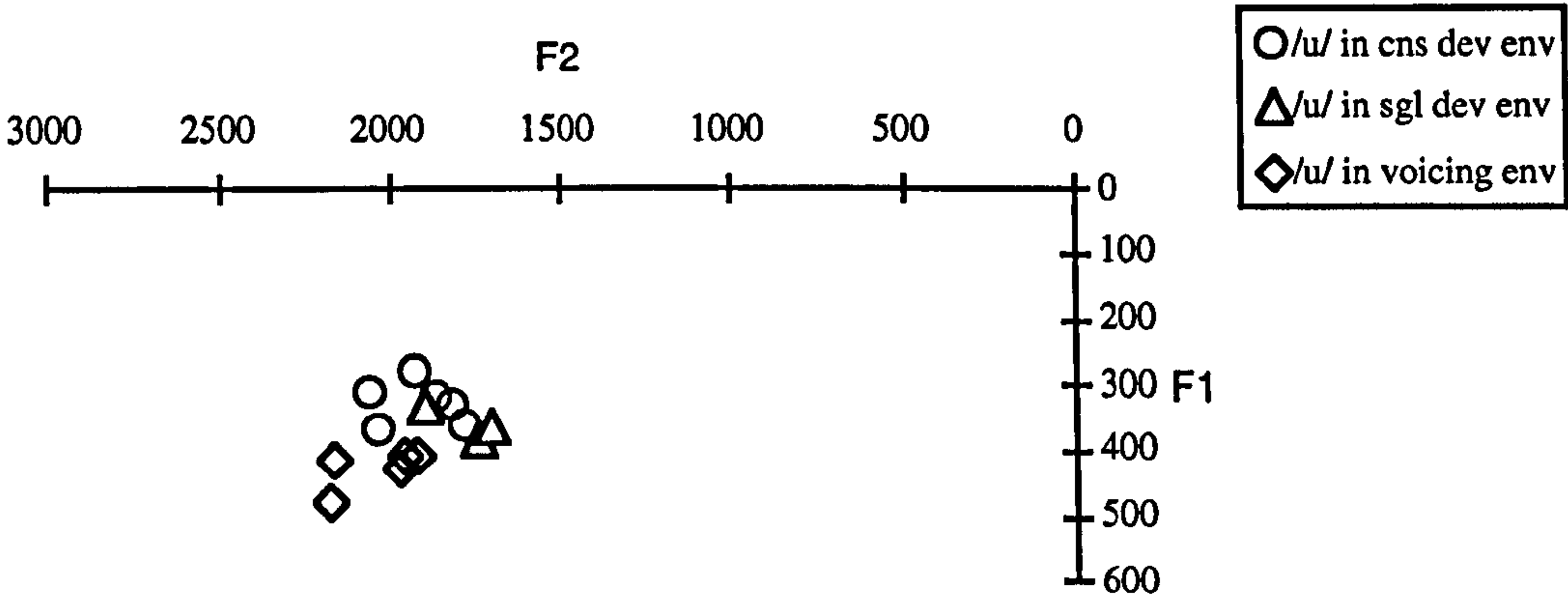
(c) A: Slow tempo /i/



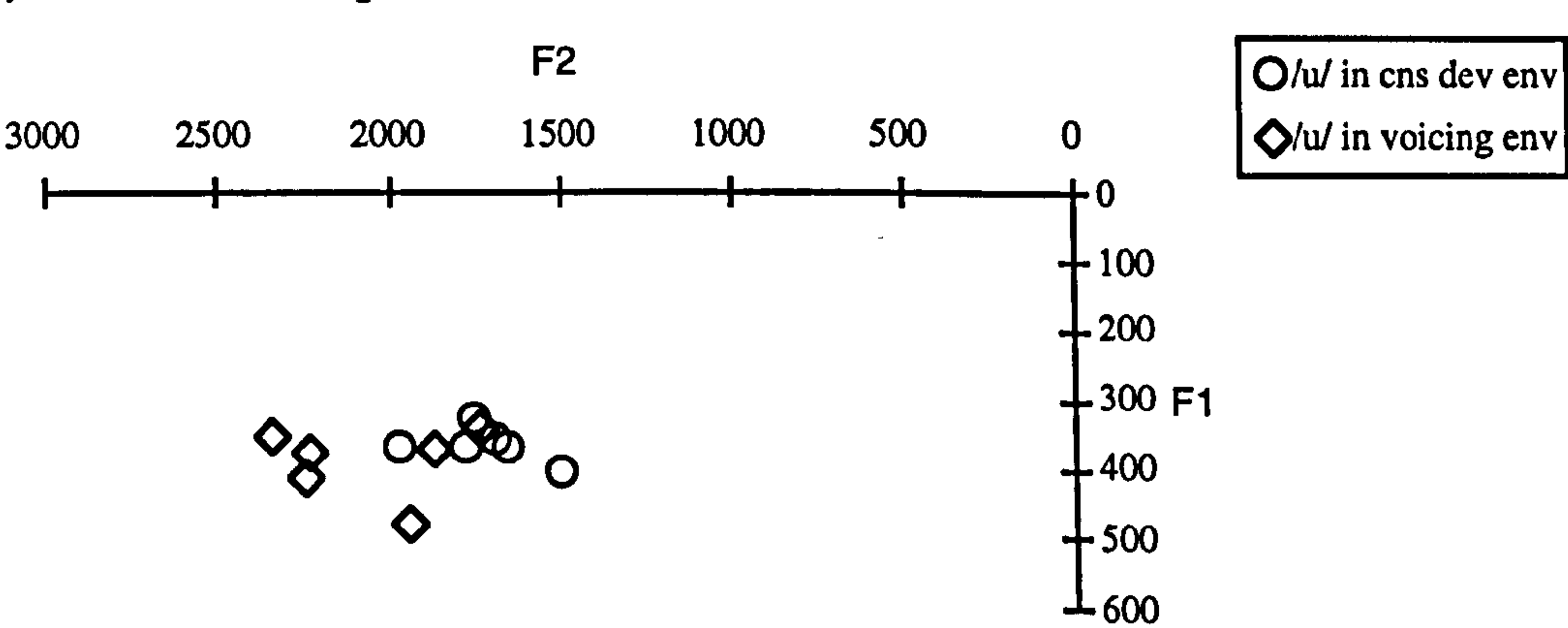
Figures 8.8a ~ 8.8c

F1 and F2 of individual tokens of /u/ in the consecutive devoicing, single devoicing and non-devoicing environments at different tempi by subject A (female)

(a) A: Fast tempo /u/



(b) A: Normal tempo /u/



(c) A: Slow tempo /u/

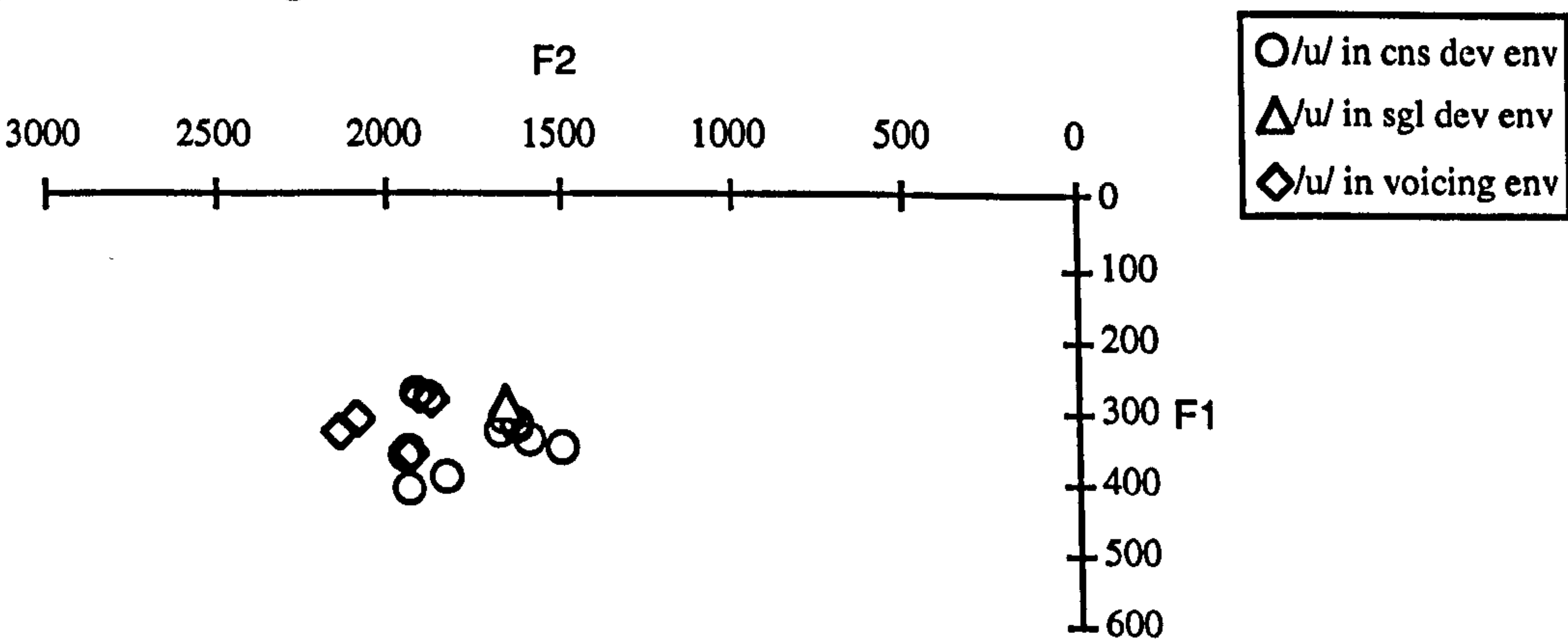
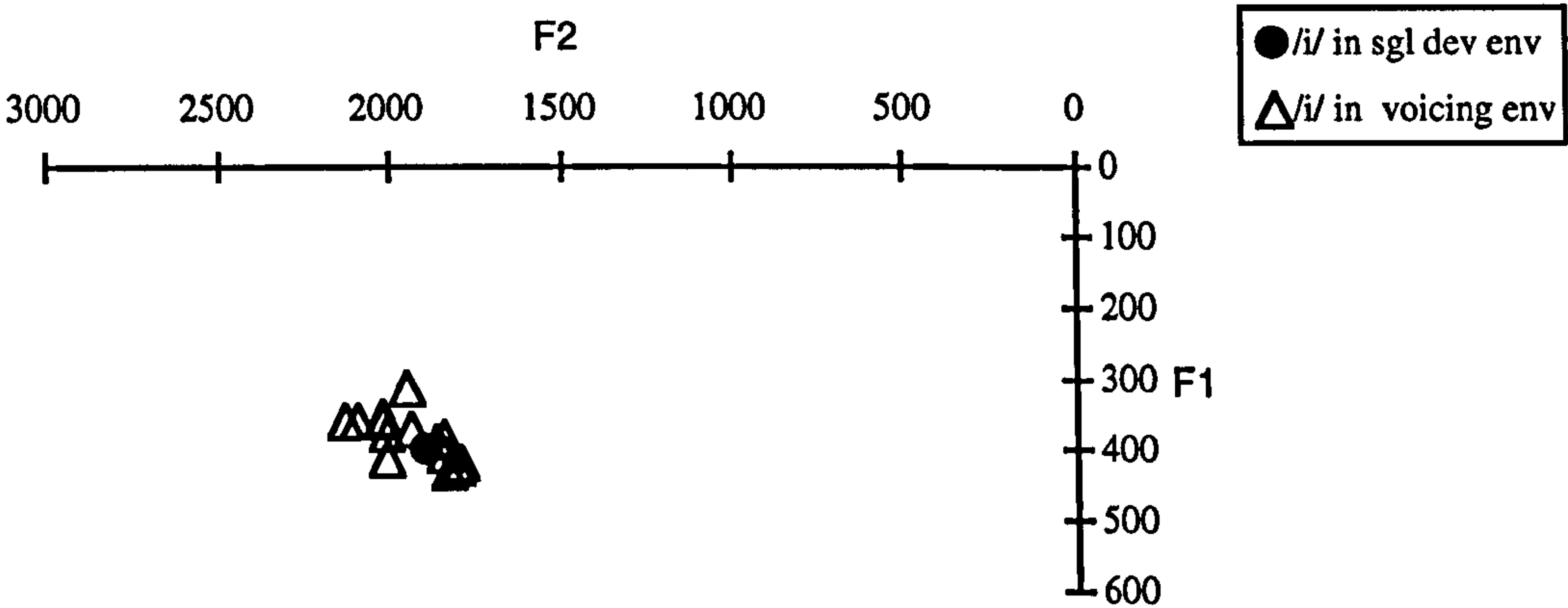


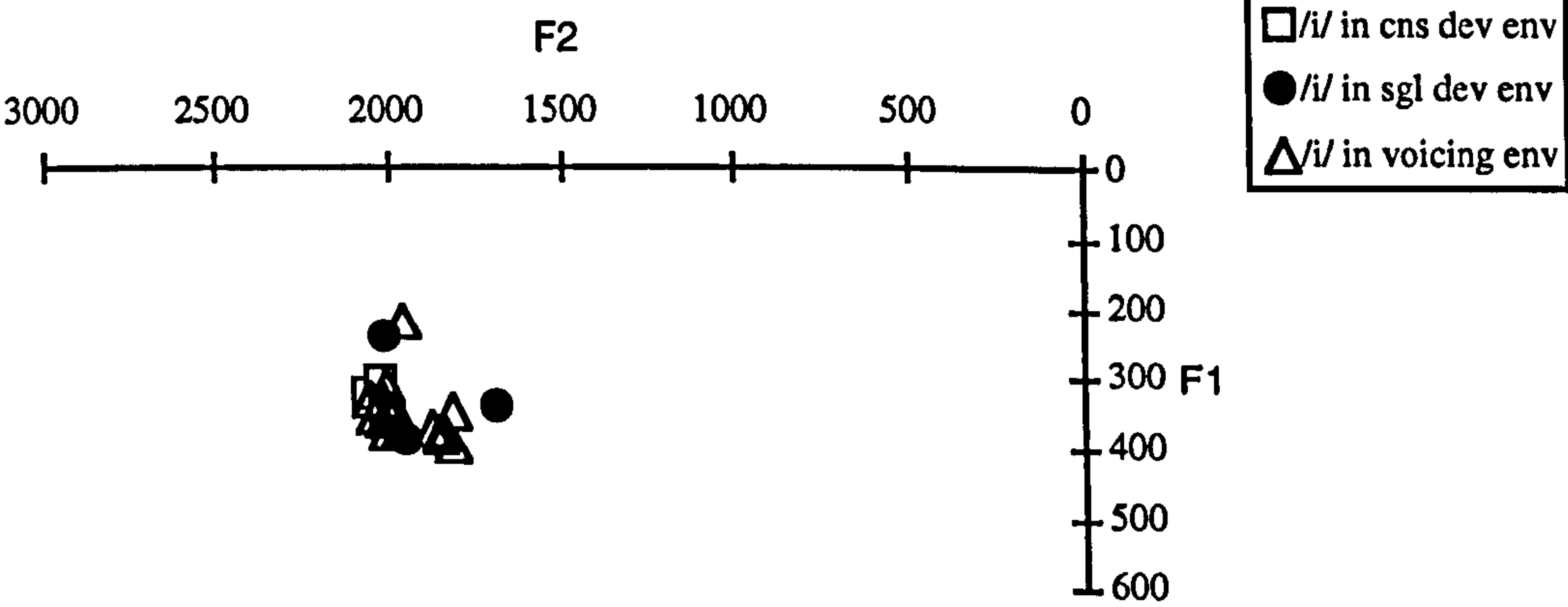
Figure 8.9a ~ 8.9c

F1 and F2 of individual tokens of /i/ in the consecutive devoicing, single devoicing and non-devoicing environments at different tempi by subject B (male)

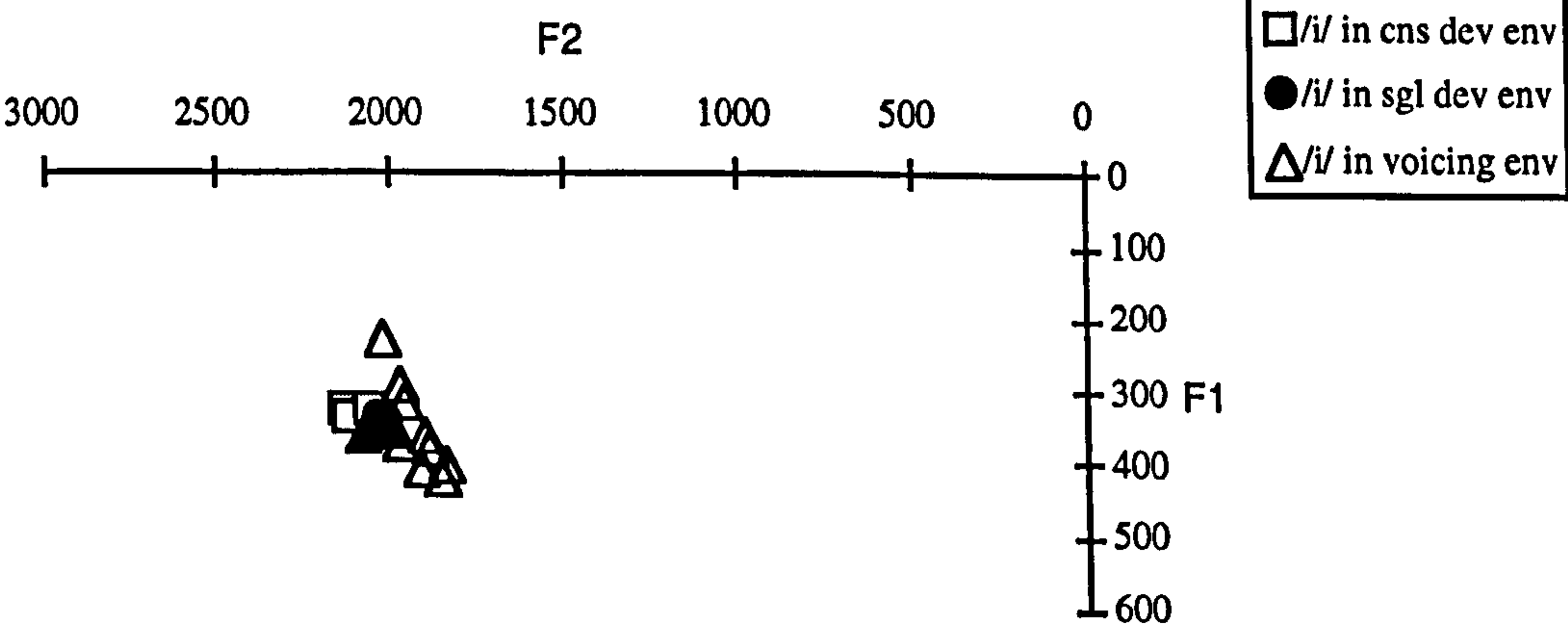
(a) B: Fast tempo /i/



(b) B: Normal tempo /i/



(c) B: Slow tempo /i/



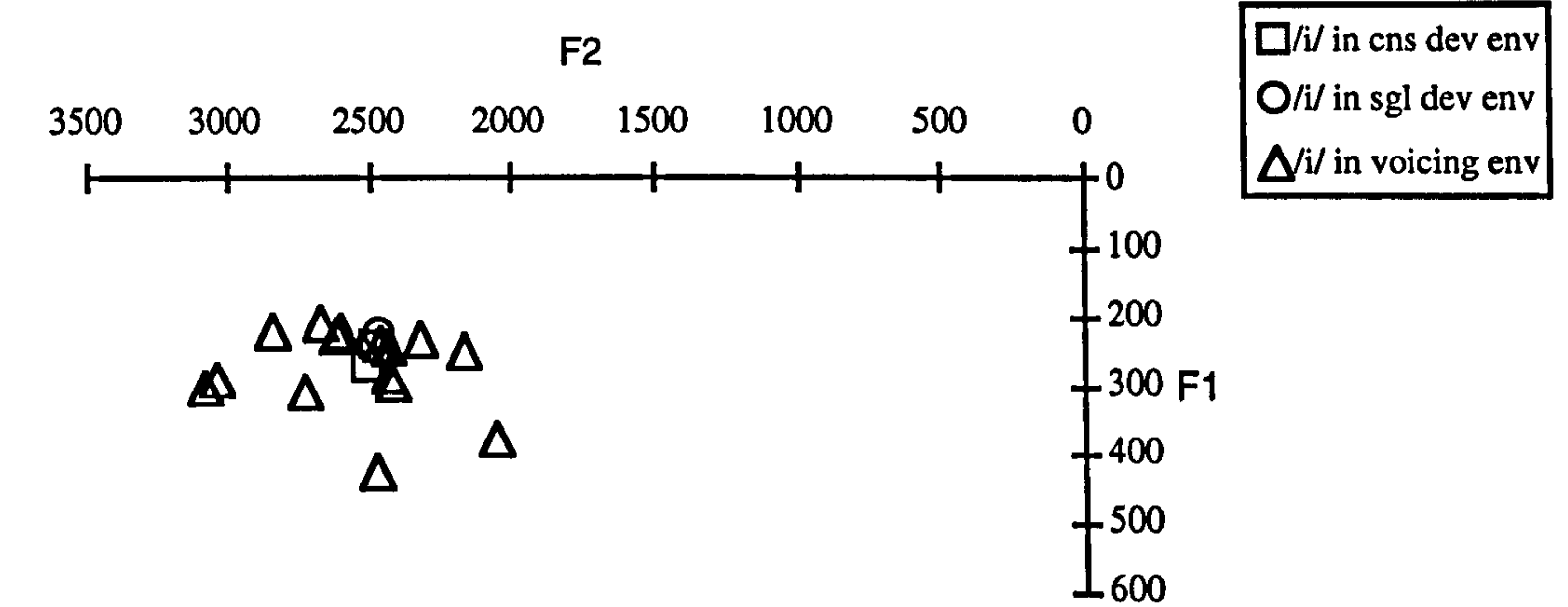




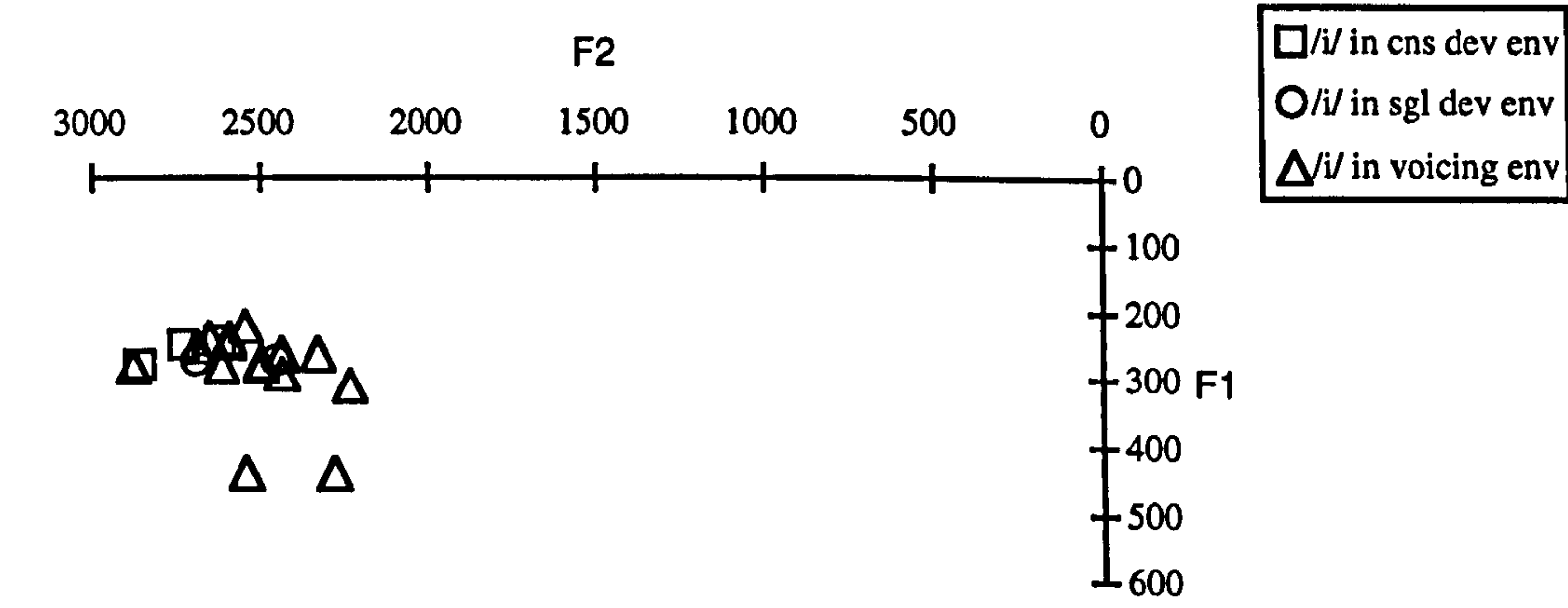
Figures 8.11a ~ 8.11c

F1 and F2 of individual tokens of /i/ in the consecutive devoicing, single devoicing and non-devoicing environments at different tempi by subject C (female)

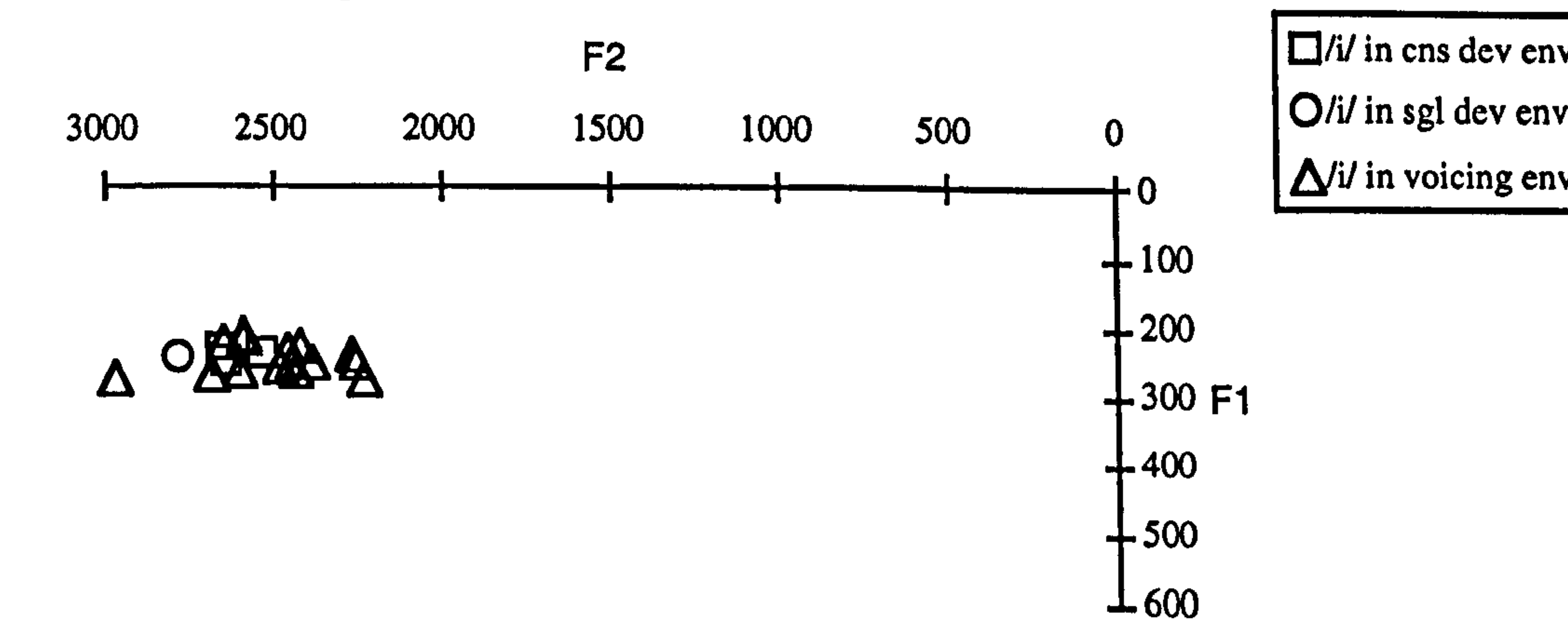
(a) C: Fast tempo /i/



(b) C: Normal tempo /i/



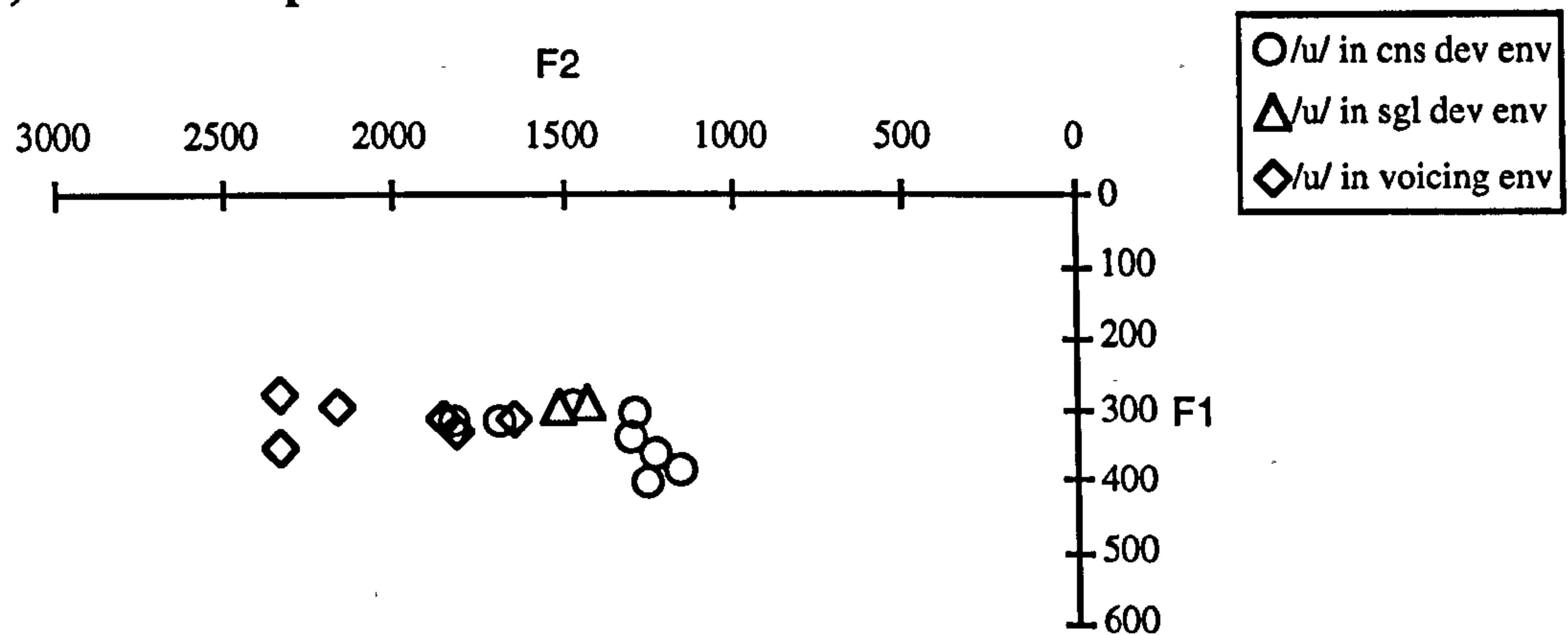
(c) C: Slow tempo /i/



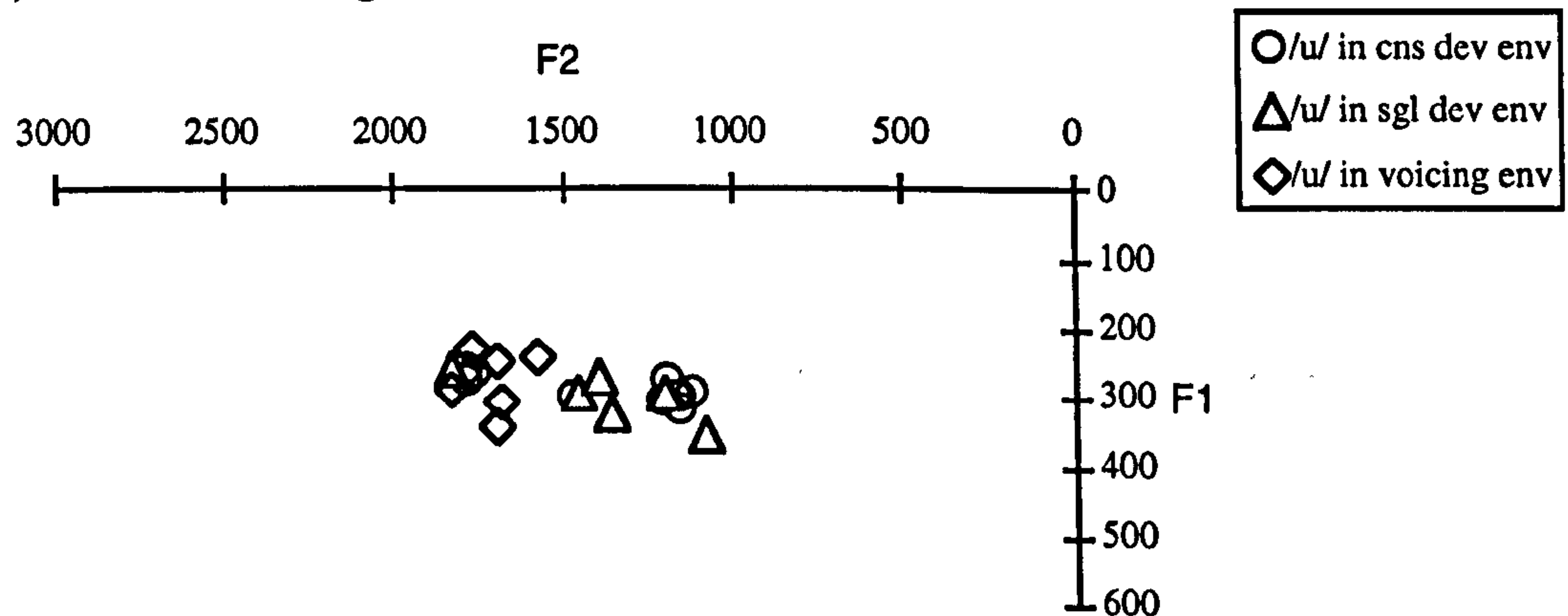
**Figures 8.12a ~ 8.12c**

**F1 and F2 of individual tokens of /u/ in the consecutive devoicing, single devoicing and non-devoicing environments at different tempi by subject C (female)**

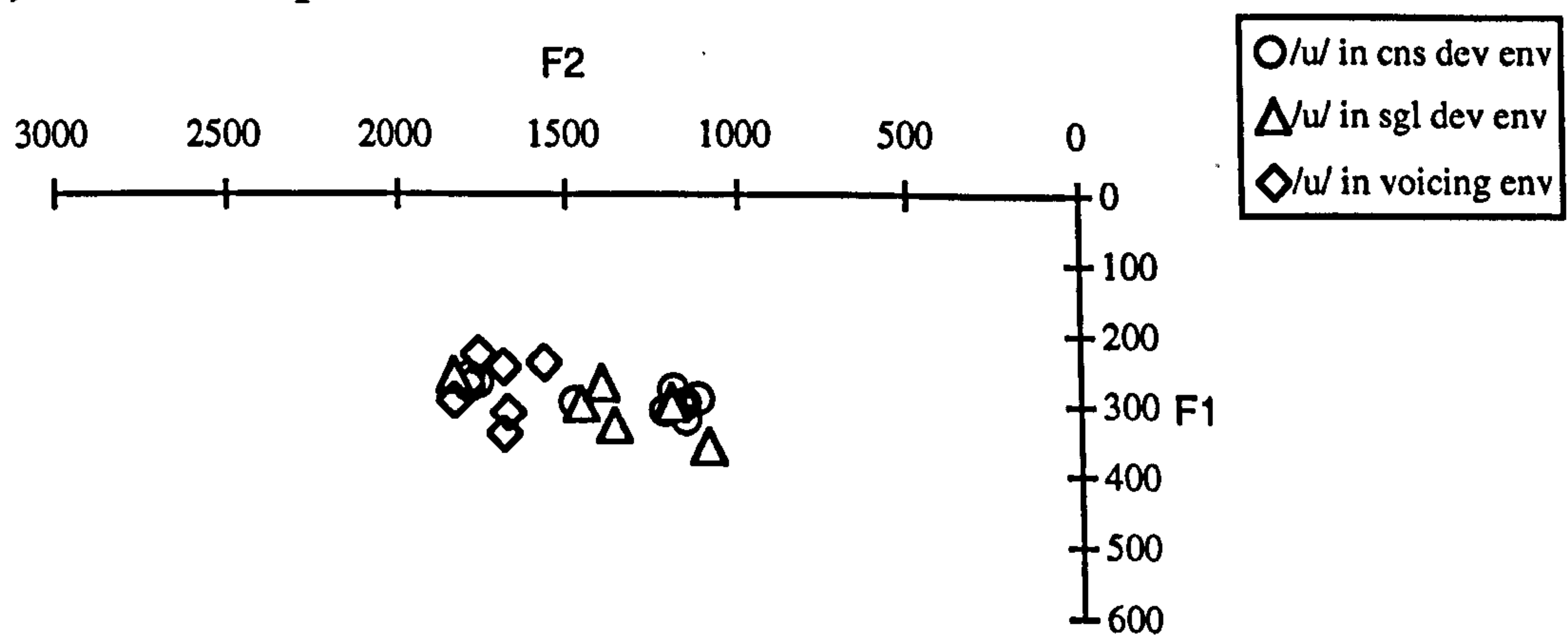
**(a) C: Fast tempo /u/**



**(b) C: Normal tempo /u/**



**(c) C: Slow tempo /u/**





Tables 8.9a ~ 8.9c

Mean and standard deviations of F1 and F2 frequencies of high vowels by devoicing conditions at three tempi by speaker A (female)

(a) Fast tempo

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	6	333	37	1913	113
/u/ in single devoicing environment	3	357	26	1782	105
/u/ in voicing environment	6	435	35	2065	122
/i/ in single devoicing environment	2	532	3	2331	98
/i/ in voicing environment	15	369	51	2346	170

(b) Normal tempo

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	6	371	24	1725	158
/u/ in voicing environment	6	380	49	2067	244
/i/ in consecutive devoicing environment	2	337	18	2550	142
/i/ in voicing environment	15	438	32	2472	140

(c) Slow tempo

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	9	351	35	1738	148
/u/ in single devoicing environment	1	(292)	N/A	(1661)	N/A
/u/ in voicing environment	6	314	30	2015	115
/i/ in consecutive devoicing environment	1	(352)	N/A	(2648)	N/A
/i/ in voicing environment	14	355	75	2462	139

**Tables 8.10a ~ 8.10c**  
**Mean and standard deviations of F1 and F2 frequencies of high vowels by devoicing conditions at three tempi by speaker B (male)**

**(a) Fast tempo**

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	4	388	21	1875	140
/u/ in single devoicing environment	2	379	50	1435	54
/u/ in voicing environment	6	377	12	1675	173
/i/ in single devoicing environment	1	(399)	N/A	(1896)	N/A
/i/ in voicing environment	15	379	33	1954	114

**(b) Normal tempo**

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	8	384	19	1658	149
/u/ in single devoicing environment	4	403	13	1384	46
/u/ in voicing environment	6	372	32	1622	181
/i/ in consecutive devoicing environment	3	329	22	2029	27
/i/ in single devoicing environment	3	326	74	1881	173
/i/ in voicing environment	15	351	43	1945	90

**(c) Slow tempo**

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	9	387	19	1477	51
/u/ in single devoicing environment	9	379	22	1306	60
/u/ in voicing environment	6	398	21	1611	139
/i/ in consecutive devoicing environment	3	331	5	2107	34
/i/ in single devoicing environment	3	347	7	2032	18
/i/ in voicing environment	15	356	47	1982	78

**Tables 8.11a ~ 8.11c**  
**Mean and standard deviations of F1 and F2 frequencies of high vowels**  
**by devoicing conditions at three tempi by speaker C (female)**

**(a) Fast tempo**

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	8	339	40	1403	238
/u/ in single devoicing environment	2	290	3	1481	60
/u/ in voicing environment	6	311	25	2032	293
/i/ in consecutive devoicing environment	2	258	21	2492	18
/i/ in single devoicing environment	2	235	10	2483	26
/i/ in voicing environment	15	273	61	2564	286

**(b) Normal tempo**

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	9	350	59	1404	298
/u/ in single devoicing environment	4	340	76	1572	252
/u/ in voicing environment	6	293	38	1948	257
/i/ in consecutive devoicing environment	3	260	20	2739	119
/i/ in single devoicing environment	2	278	1	2580	170
/i/ in voicing environment	15	289	66	2529	165

**(c) Slow tempo**

Vowel	No.	F1 (Hertz)		F2 (Hertz)	
		mean	SD	mean	SD
/u/ in consecutive devoicing environment	10	292	19	1381	292
/u/ in single devoicing environment	7	300	33	1365	247
/u/ in voicing environment	6	277	44	1711	87
/i/ in consecutive devoicing environment	3	246	8	2609	64
/i/ in single devoicing environment	1	(249)	N/A	(2779)	N/A
/i/ in voicing environment	15	250	21	2498	191



It was not easy to make a good comparison of F1 and F2 between devoicing and voicing environments since the number of samples was limited; there were not many voiced high vowels in the devoicing environment, and phonetic environments for each vowel differed enormously. However, several tendencies were observed.

As mentioned earlier (Section 8.1), it is widely acknowledged that Japanese vowels do not have the characteristics of obvious centralisation and durational reduction related to accentuation (Keating and Huffman, 1984). However, for all three speakers, the high vowels /i/ and /u/ seemed to be neutralised especially when they were in non-devoicing environments. /u/ had very high F2, sometimes almost as high as F2 of /i/, while /i/ had low F1 frequencies, almost as low as F1 of /u/. The Japanese vowels /i/ and /u/ are usually phonetically described as [i] and [ɯ] respectively, but the quality of /u/ here was more central like [ɨ] or [ʊ]. This may have been due to the small number of samples in the experiment.

Multivariate analyses of variance (MANOVA) were performed taking F1 and F2 values as dependent variables, and devoicing environment and speaking tempo as factors, for each vowel by individual subjects. MANOVA results were based on Wilks' Lambda. The results are presented in Tables 8.12(a, b) ~ 8.17(a, b).

Tables 8.12a and 8.12b

ANOVA and MANOVA tables for vowel /i/ by subject A

(a) ANOVA table of F1 and F2

	F1			F2		
	df	F-value	p-value	df	F-value	p-value
(a)Environment	2	8.016	p <.001	2	.957	n.s.
(b) Tempo	2	.591	n.s.	2	3.405	p <.05
(a) x (b)	1	.012	n.s.	1	.312	n.s.
Residual	43			43		

(b) MANOVA Table

Effect	F-value	p-value
(a) Environment	5.187	p <.001
(b) Tempo	1.729	n.s.
(a) x (b)	.253	n.s.

Tables 8.13a and 8.13b

ANOVA and MANOVA tables for vowel /u/ by subject A

(a) ANOVA table of F1 and F2

	F1			F2		
	df	F-value	p-value	df	F-value	p-value
(a)Environment	2	.444	n.s.	2	15.856	p <.0001
(b) Tempo	2	1.444	n.s.	2	1.589	n.s.
(a) x (b)	3	.864	n.s.	3	.838	n.s.
Residual	35			35		

(b) MANOVA Table

Effect	F-value	p-value
(a) Environment	6.569	p <.001
(b) Tempo	1.4659	n.s.
(a) x (b)	.769	n.s.

Tables 8.14a and 8.14b

ANOVA and MANOVA tables for vowel /i/ by subject B

(a) ANOVA table of F1 and F2

	F1			F2		
	df	F-value	p-value	df	F-value	p-value
(a)Environment	2	.402	n.s.	2	3.021	n.s.
(b) Tempo	2	1.949	n.s.	2	2.943	n.s.
(a) x (b)	3	.290	n.s.	3	.716	n.s.
Residual	50			50		

(b) MANOVA Table

Effect	F-value	p-value
(a) Environment	1.534	n.s.
(b) Tempo	2.975	p <.05
(a) x (b)	.642	n.s.

Tables 8.15a and 8.15b

ANOVA and MANOVA tables for vowel /u/ by subject B

(a) ANOVA table of F1 and F2

	F1			F2		
	df	F-value	p-value	df	F-value	p-value
(a)Environment	2	.679	n.s.	2	13.309	p <.0001
(b) Tempo	2	1.992	n.s.	2	7.833	p <.001
(a) x (b)	4	2.903	p <.05	4	2.773	p <.05
Residual	44			44		

(b) MANOVA Table

Effect	F-value	p-value
(a) Environment	7.210	p <.0001.
(b) Tempo	3.877	p <.01
(a) x (b)	3.232	p <.01



Tables 8.16a and 8.16b

ANOVA and MANOVA tables for vowel /i/ by subject C

(a) ANOVA table of F1 and F2

	F1			F2		
	df	F-value	p-value	df	F-value	p-value
(a)Environment	2	.533	n.s.	2	.768	n.s.
(b) Tempo	2	.682	n.s.	2	.764	n.s.
(a) x (b)	4	.188	n.s.	4	.906	n.s.
Residual	49			49		

(b) MANOVA Table

Effect	F-value	p-value
(a) Environment	.541	n.s.
(b) Tempo	.762	n.s.
(a) x (b)	.562	n.s.

Tables 8.17a and 8.17b

ANOVA and MANOVA tables for vowel /u/ by subject C

(a) ANOVA table of F1 and F2

	F1			F2		
	df	F-value	p-value	df	F-value	p-value
(a)Environment	2	3.423	p <.05	2	21.447	p <.0001
(b) Tempo	2	4.134	p <.05	2	2.245	n.s.
(a) x (b)	4	1.052	n.s.	4	.749	n.s.
Residual	49			49		

(b) MANOVA Table

Effect	F-value	p-value
(a) Environment	9.112	p <.0001
(b) Tempo	4.554	p <.01
(a) x (b)	.947	n.s.

Statistical results showed that the effect of factors varied for each factor, subject and vowel. For subject A, the effect of environment was significant on F1 for /i/ and F2 for /u/, and the effect of tempo was significant on only F2 for /i/. MANOVA results suggested that there was a significant overall effect of the devoicing environment on F1 and F2 values of both /i/ and /u/. As for subject B, no significant effect was found on either formant of /i/, but the effect of both factors on the F2 value of /u/ was significant. There was also a significant interaction between the environment and tempo on F2 of /u/. MANOVA results suggested that the overall effect of tempo on dependent variables was significant for /i/, and the overall effects of both factors were significant and significant interaction between the environment and tempo was found for /u/. As for subject C, the effect of the environment on F1 and F2 values of /u/ was found to be significant, but the effect of tempo was significant only on F2 values of /u/. MANOVA results suggested, however, that overall effects of both factors, the environment and tempo, were significant. However, there were no significant interactions.

Some statistical results showed significant effects of the devoicing environment and speaking tempo on F1 and/or F2 values of both vowels, and also an overall significant effect of both factors on F1 and F2. However, the effects were not consistent. Neither of the factors significantly influenced either dependent variable consistently for both vowels and all subjects. None of the MANOVA results showed consistent effects, either. Some of the significant levels were very high. However, the results obtained might have been caused by inconsistent experimental environments.

Keating and Huffman (1984) measured the F1 and F2 of Japanese vowels from readings of word lists and prose text. Table 8.18 shows the mean F1 and F2 frequencies of Japanese high vowels sampled from a text reading obtained by Keating and Huffman.

**Table 8.18 Means in Hertz of F1 and F2 for high vowels in prose  
(all male speakers)**

Keating and Huffman (1984: 199)

vowel	formant	speaker 1	speaker 2	speaker 3	speaker 4	speaker 5	speaker 6	speaker 7
/i/	F1	347	348	350	364	379	357	371
	F2	2001	1753	1843	2120	1973	2140	1851
/u/	F1	386	398	403	410	423	402	413
	F2	1493	1352	1345	1482	1504	1402	1357

The number of tokens per vowel of the results by Keating and Huffman varied from 59 to 100, but the figures were probably more accurate than the figures of my results whose number of tokens per vowel varied from 1 to 15. Their results for formant frequencies for all 5 vowels /i/, /e/, /a/, /o/ and /u/ in word lists did not vary very much, but it was in continuous prose that the formant frequencies showed a lot of variation.

Formant frequencies of high vowels in my experiment showed a large variation as expected (see Figures 8.7a ~ 8.7c, 8.8a ~ 8.8c, 8.9a ~ 8.9c, 8.10a ~ 8.10c, 8.11a ~ 8.11c and 8.12a ~ 8.12c). In fact, the vowels in voicing environments showed more variation than the vowels in devoicing environments. It was probably because the vowels in voicing environments had more tokens, and their phonetic environments varied more than the vowels in devoicing environments, which were always surrounded by voiceless obstruents.

The intensity ratios in Section 8.4.3 showed that the vowels in single devoicing sites were weakened more than the vowels in consecutive devoicing sites for all the speakers. If there is a positive correlation between vowel intensity and vowel quality, the lower the intensity, the more centralised a vowel, as reported for languages such as English and German. As far as the quality of vowels is concerned, however, there seemed to be no consistent effect of the devoicing environment on vowel qualities as between single devoicing sites and consecutive devoicing sites.



Both vowels seemed to be centralised more at the faster tempi than at the slower tempi, but very little difference in quality was found in relation to vowel intensity. The quality of vowels seems to be affected more by speech rate and speech style.

The results of this experiment in Japanese showed that vowels in devoicing environments were weakened in intensity, and shortened in duration, but the quality did not necessarily change very much. The result differs a great deal from the vowel reduction process in other languages reported in Chapter 2. The mora-based rhythmic structure of Japanese may require different vowel reduction processes.

### 8.5 Vowel devoicing processes

As described in Chapter 2, devoicing in many languages mainly involves high vowels and schwa which are intrinsically short in duration, but the environments are not restricted to being adjacent to voiceless consonants. The typical environment for devoicing is unstressed syllables, regardless of adjacent sounds, and devoicing is treated as an optional fast speech rule in a conversational style. This process is particularly common in languages in which the function of stress is particularly important, and it involves changes of vowel quality: vowels become centralised in unaccented syllables, then are reduced to a schwa, and are eventually devoiced, or in extreme cases, such as in very fast speech, they are deleted.

In Japanese, devoicing involves high vowels which are intrinsically short in duration, but are full vowels. Japanese does not have an obvious process of vowel reduction in relation to stress or intonation, although a slight vowel centralisation occurs in reading style (Keating and Huffman, 1984). A crucial factor is that vowel devoicing in Japanese is not an optional fast speech rule; it occurs at slow or normal tempi. The experimental results in Chapters 7 and 8 showed that the Japanese vowel devoicing processes involve the weakening of the intensity of vowels in devoicing environments as well as a phonatory change and durational reduction. The whole devoicing process is a vowel weakening process as in other languages. The only

difference is that the application of each process is different in Japanese from in other languages. How does the devoicing process operate?

As described in Chapter 2, Section 2.4.4, Jun (1993), Jun and Beckman (1993 and 1994) studied vowel devoicing in Korean with relation to glottal gestural overlap and suggested that the same process applied to Japanese vowel devoicing. In Korean, high vowel devoicing is largely dependent on the glottal width of preceding consonants. Devoicing frequency is higher when a vowel is preceded by stops showing a large glottal-opening gesture, and lower when preceded by stops with a much smaller glottal opening gesture. Therefore, Jun and Beckman concluded that vowel devoicing in Korean was the result of glottal gestural overlap of neighbouring vowel and consonants.

Jun and Beckman applied the same gestural-overlap theory to Japanese vowel devoicing based on Yoshioka's study of the posterior cricoarytenoid (PCA: abducting muscles) and interarytenoid (INT: adducting muscles) in Japanese (Yoshioka, 1981, also see Chapter 4, Section 4.3). Yoshioka's electromyographic data showed the blending gesture of the abductor (PCA) and adductor (INT) when a vowel was devoiced between voiceless consonants. When a vowel was voiced between voiceless consonants, there were two separate movements of the PCA for the voiceless vowels, and the INT showed a movement for the vowel. However, when a vowel was devoiced, the PCA showed only one big movement between the voiceless consonants and the movement of the adductor was suppressed between the voiceless consonants.

Vowel devoicing in Japanese typically involves high vowels which are inherently short in duration, and between voiceless consonants or a pause for both of which the glottis is wide open. This is an ideal environment for the glottal gestures of both voiceless consonants and the high vowels to blend with each other. Therefore, Jun and Beckman proposed that vowel devoicing in Japanese, as in Korean, was the result of glottal gestural-overlap.

If vowel devoicing in Japanese is also the result of blending glottal gestures, it should occur more at faster tempi than slower tempi, where each articulator, including the glottal articulators, achieves its target properly, and does not overlap its gesture. My experimental results in Chapters 6 and 8 suggested that vowel devoicing rates did not always vary very much depending on speaking tempo, but the difference of tempo was manifested in the vowel intensity, i.e. devoiceable vowels were weakened more at faster tempi.

Vowels in devoicing environments could be voiced, but voiced devoiceable vowels were weaker than ordinary vowels. This may imply that it is natural for high vowels to undergo devoicing, rather than remaining voiced between voiceless consonants. This is supported by the fact that voiced devoiceable vowels were generally weakened more at faster tempi and also that vowels were almost always devoiced so long as there was only one devoiceable vowel in the word. The results confirmed that vowel devoicing in Japanese is part of the vowel weakening process. The vowel weakening process in Japanese does not work in the same way as vowel reduction in stress-timed languages; vowel weakening is realised as a weakening of intensity rather than a reduction of duration or.

A problem is that, as mentioned in Chapter 4, some physiological studies reported positive opening of the glottis during the production of voiceless vowels (Sawashima, 1971), and especially the vowel devoicing following fricatives showed complete merging of the gestures of laryngeal muscle activities (Simada et al, 1991). The laryngeal muscular activities for voiceless vowels were clearly different from other coarticulatory sequences, such as voicing of an intervocalic /h/. This implies that high vowel devoicing may have originally been a result of glottal gestural overlap with adjacent voiceless consonants as has been suggested in Yoshioka's results, but the process has become established over years and now has become a compulsory process.

Another factor relating to the gestural overlap analysis is that not all devoiceable vowels undergo the devoicing process in the consecutive devoicing environment, but



some devoiced vowels remain voiced (see Chapters 5 and 6). In the consecutive devoicing environment, blocking factors became effective preventing some devoiced vowels from becoming voiceless and voiced devoiced vowels were not weakened in intensity (see Chapter 5 and Section 8.4 in this chapter). Japanese vowel devoicing seems to involve other factors, as well as glottal gesture. Moreover, voiced devoiced vowels in consecutive devoicing environments were not weakened. If vowels were devoiced consecutively over two morae, it would create a series of consonant clusters on the surface level, which is not a favoured sound structure in Japanese. The vowels can be blended into adjacent voiceless consonants, but there are constraints which prevent some vowels from undergoing the devoicing process, so that the preferred syllable structure is retained on the surface level.

One possible solution might be found in the phonological analysis of the vowel deletion process in French. For example, in French, which is a syllable-timed language, vowels are not normally reduced and always preserve their full quality, and vowel devoicing is not the consequence of centralisation or reduction of a full vowel. The only vowel which can be reduced is [œ], which is a fully-fledged vowel and often described as schwa, under certain conditions. When this vowel is reduced, it is reduced to nothing, i.e. deleted (Tranel, 1987b: 86).

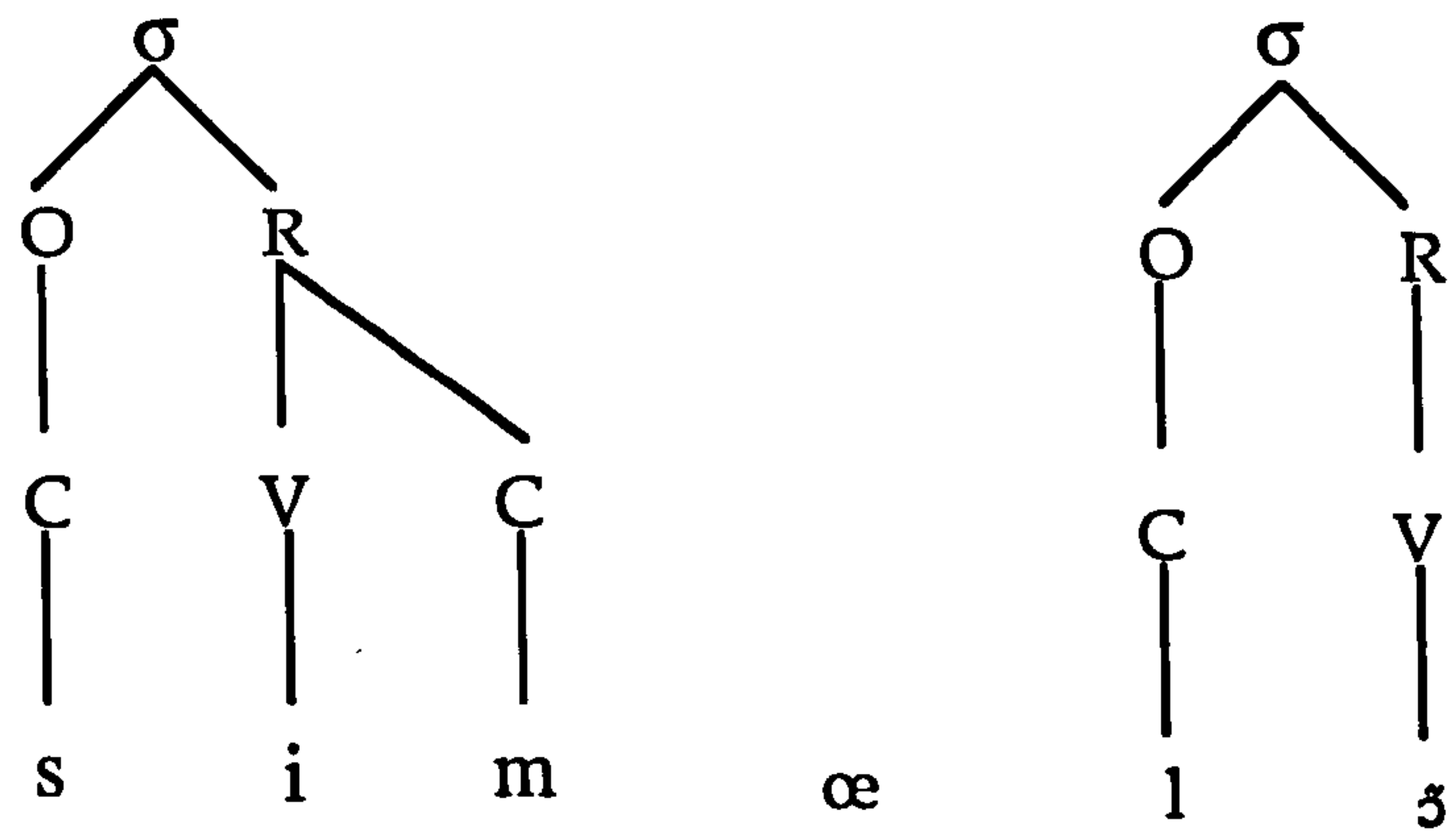
Tranel (1987a) proposed that the phonetic realisation of French schwa was always predictable because the process was constrained by syllable structure. He analysed the deletion in the framework of autosegmental phonology, and proposed that schwa ([œ]) could be deleted only if the remaining consonants were syllabified into an unoccupied coda of the preceding syllable. For example, the schwa in *six melons* /simœlɔ̃/ 'six melons' can be deleted as [simlɔ̃], but not in *sept melons* /setmœlɔ̃/ 'seven melons' as \*[setmlɔ̃]. The processes cannot be explained only in terms of French phonotactics. Morin (1978) listed examples of consonant clusters with an obligatory schwa and without schwa.

(3) (from Morin, 1978: 111)

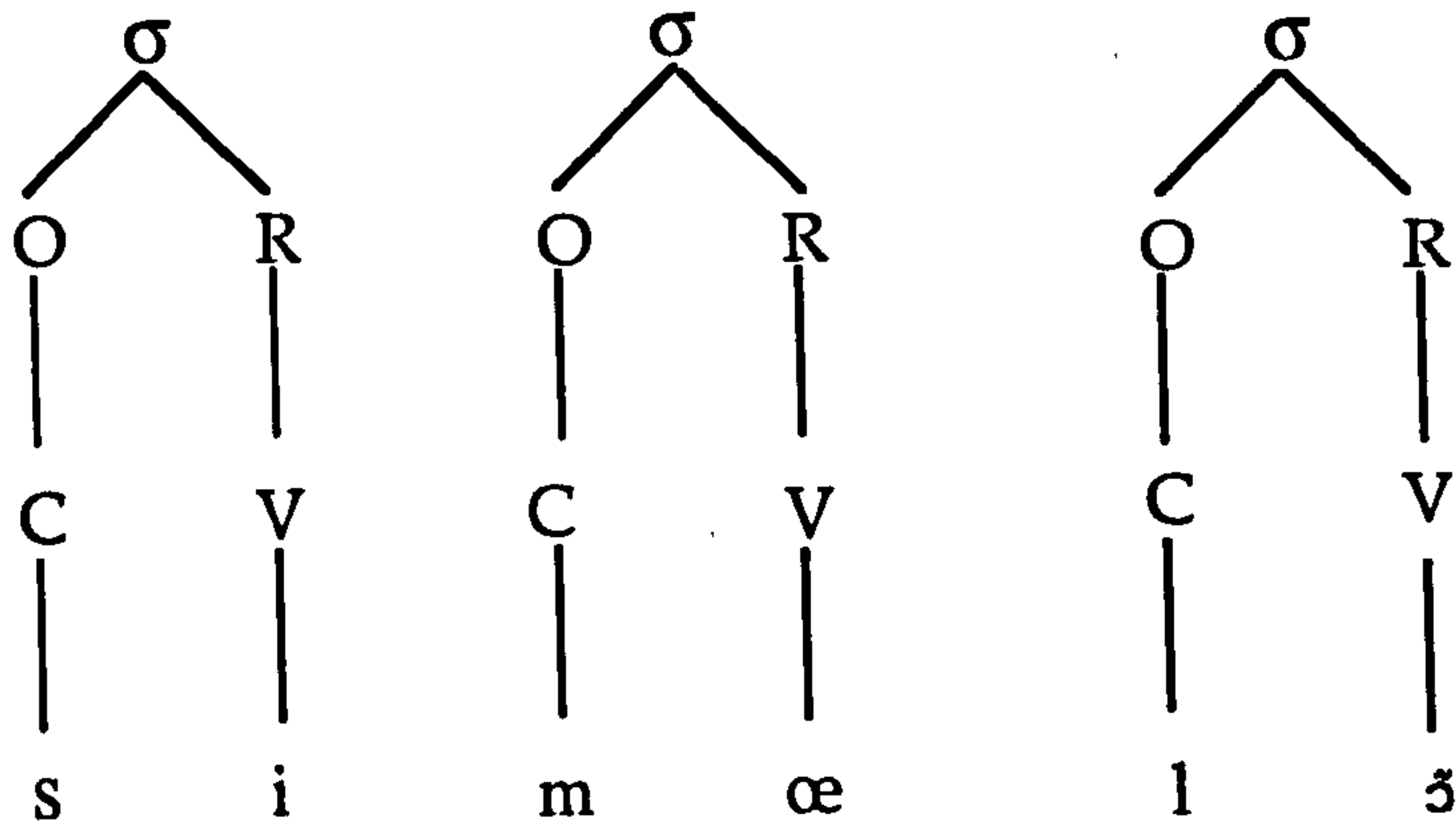
clusters	obligatory schwa			without schwa		
/rdr/	<i>bordereau</i>	[bɔ̃ʁdœʁo]	'note, slip'	<i>perdrix</i>	[pɛʁdʁi]	'partridge'
/rʃ/	<i>derechef</i>	[dœʁœʃɛf]	'once more'	<i>torchon</i>	[tɔʁʃɔ̃]	'cloth'
/skl/	<i>squelette</i>	[skœlɛt]	'skeleton'	<i>sclérose</i>	[sklɛʁoz]	'sclerosis'
/ps/	<i>dépecer</i>	[depœse]	'to cut up'	<i>éclipser</i>	[eklipse]	'to eclipse'

Tranel (1987a) proposed an analysis, with schwa as a floating vowel. He adopted a CV skeleton tier and a segmental tier. If a consonant preceding schwa can be syllabified in the coda position of a preceding syllable, as long as the new consonant sequences satisfy French phonotactics, the schwa is not required and thus can be realised as zero, i.e. deleted. On the other hand, if the consonant preceding schwa cannot be syllabified with the preceding syllable, the consonant needs the assistance of the schwa to be syllabified as an independent syllable. Therefore, for /simœlʒ/, both (4a) and (4b) are acceptable.

(4a) N.B. "σ" denotes a syllable, "O" denotes onset, and "R" denotes rhyme.

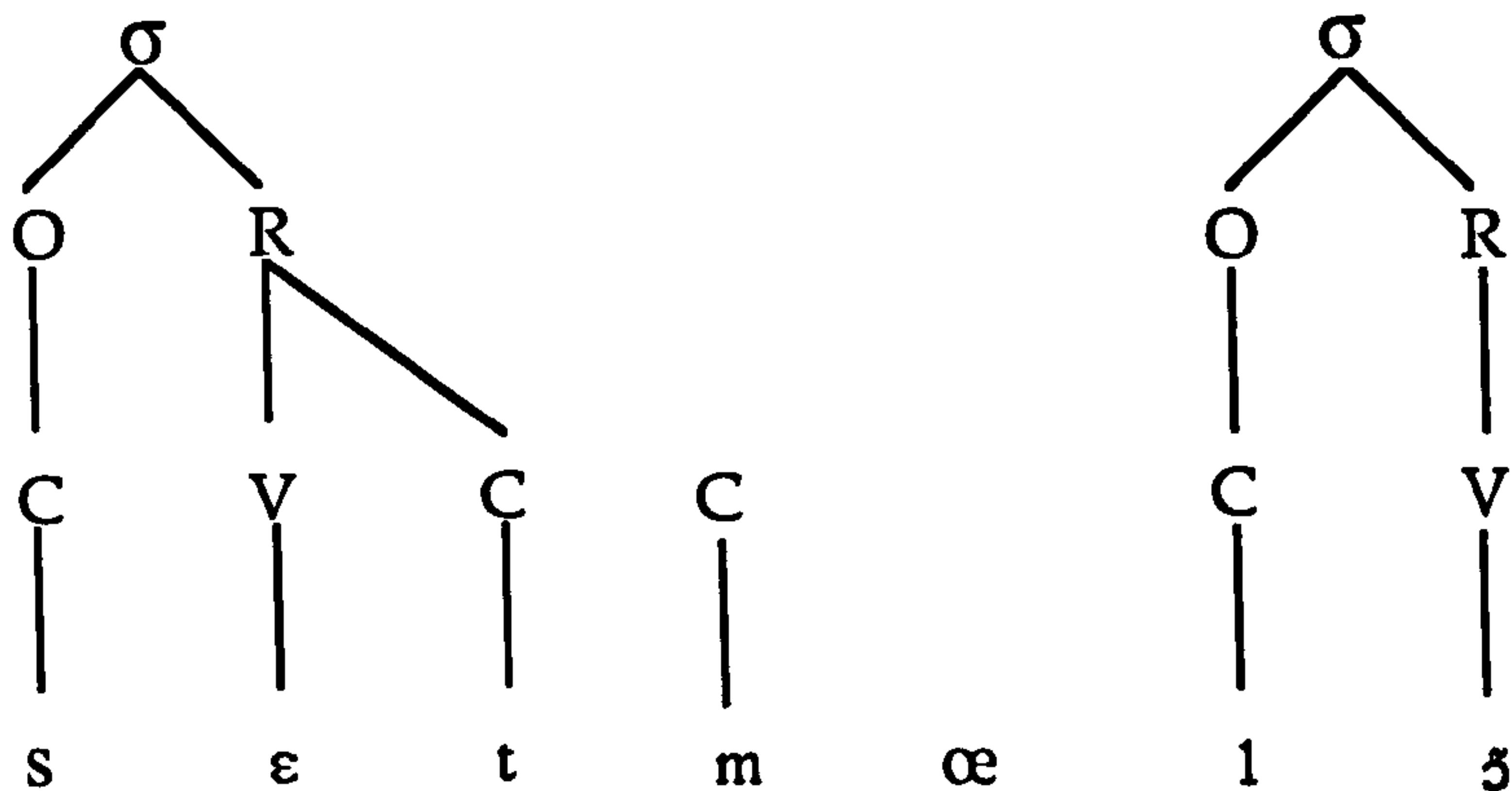


(4b)



However, in the case of /sɛtmœlɔ̃/, the schwa is obligatory because the preceding /m/ cannot be syllabified in the coda position of the preceding syllable, as shown in \*(5).

\*(5)



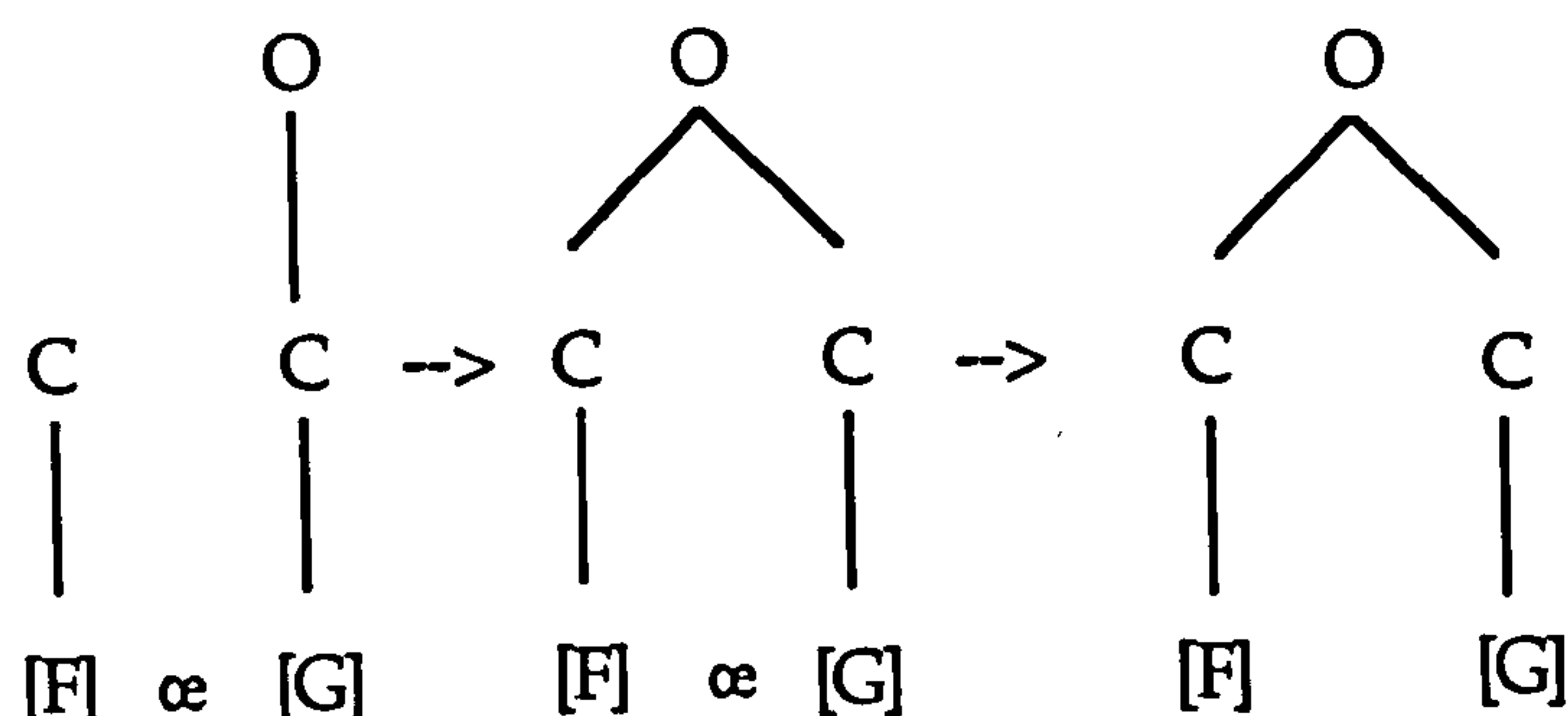
Therefore, schwa can also be deleted in *trois chemises* [tʁwɑʃ(œ)miz] 'three shirts', *la fenêtre* [laf(œ)netʁ] 'the window', and *des petits chats* [dɛp(œ)tifa] 'small cats' in all of which the preceding consonants can be syllabified to the preceding syllable, even if the schwa is deleted. However, schwa cannot be deleted in *une belle chemise* [ynbɛlfœmiz] 'a nice shirt', *une fenêtre* [ynfœnetʁ] 'a window' and *sept petits chats* [sɛtpœtifa] 'seven small cats' because the preceding consonants cannot be syllabified to the preceding syllable.



In word initial position, such as *je pars* [ʒpɑʁ] 'I'm leaving', *ce cartable* [skɑʁtable] 'this satchel', *te fais pas de bile* [tʁɛpɑdbil] 'don't worry', different rule called 'onset accretion across schwa' is applied as shown in (6).

(6) Onset accretion across schwa

N.B. [F] and [G] denote feature matrixes.

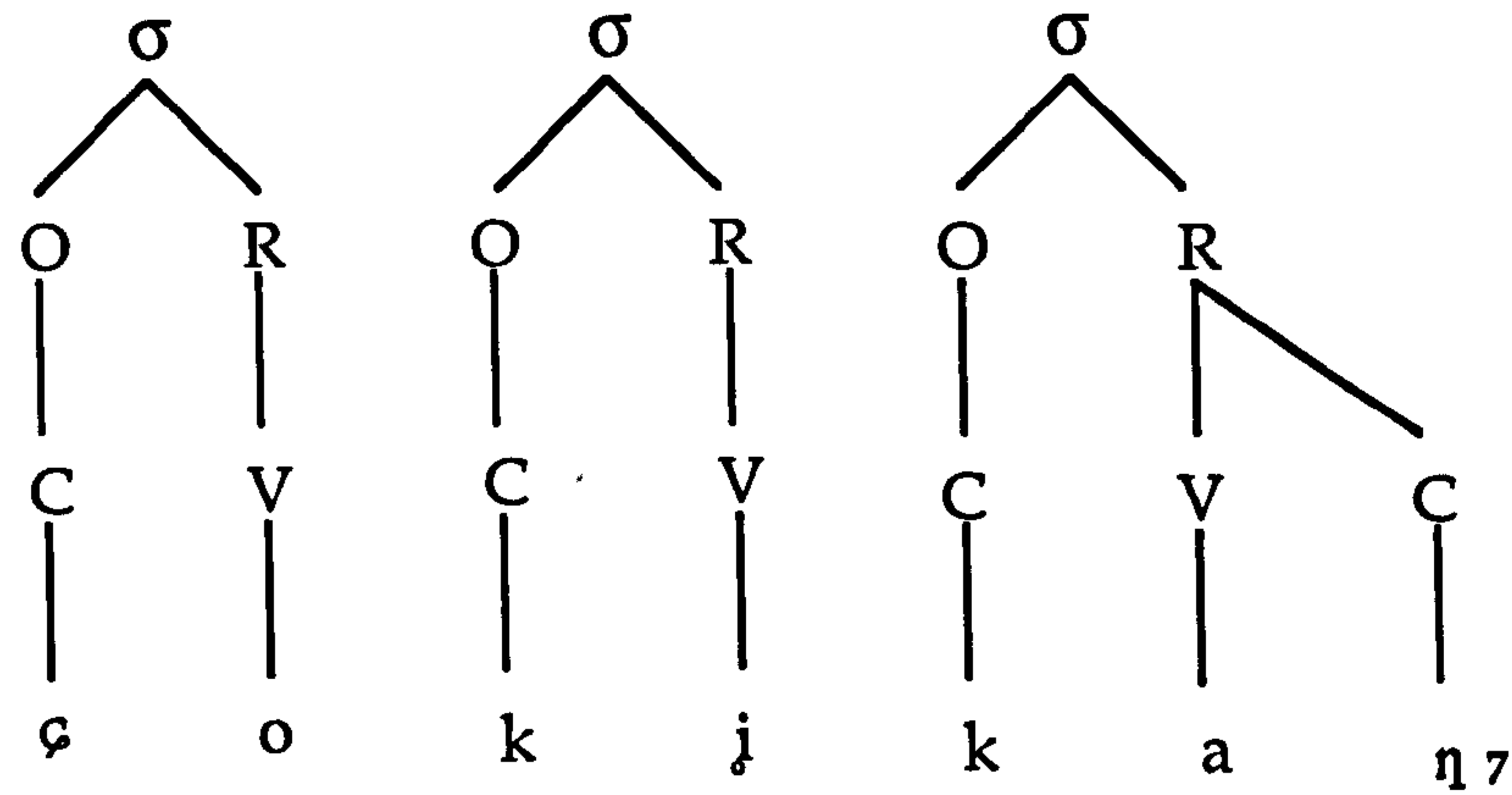


Onset accretion may be applied across schwa, as long as the resulting consonant clusters obey some criterion of pronounceability, definable in terms of the sonority interval permitted between consonants in a syllable onset (Tranel, 1987a: 852). It is always possible for the schwa in this position to be pronounced without sounding unnatural. However, this rule cannot be applied in word internal position, so as to avoid the generation of word-internal trilateral consonant clusters from underlying /CCəC/ sequences. If schwa is available to break up the cluster, it must be realised. For instance, *marguerite* /mɑʁgɛrit/ 'daisy' is always [mɑʁgɛrit] not \*[mɑʁgɛrit], and *margrave* /mɑʁɡrɑv/ 'margrave' is always [mɑʁɡrɑv] not \*[mɑʁgɛrɑv].

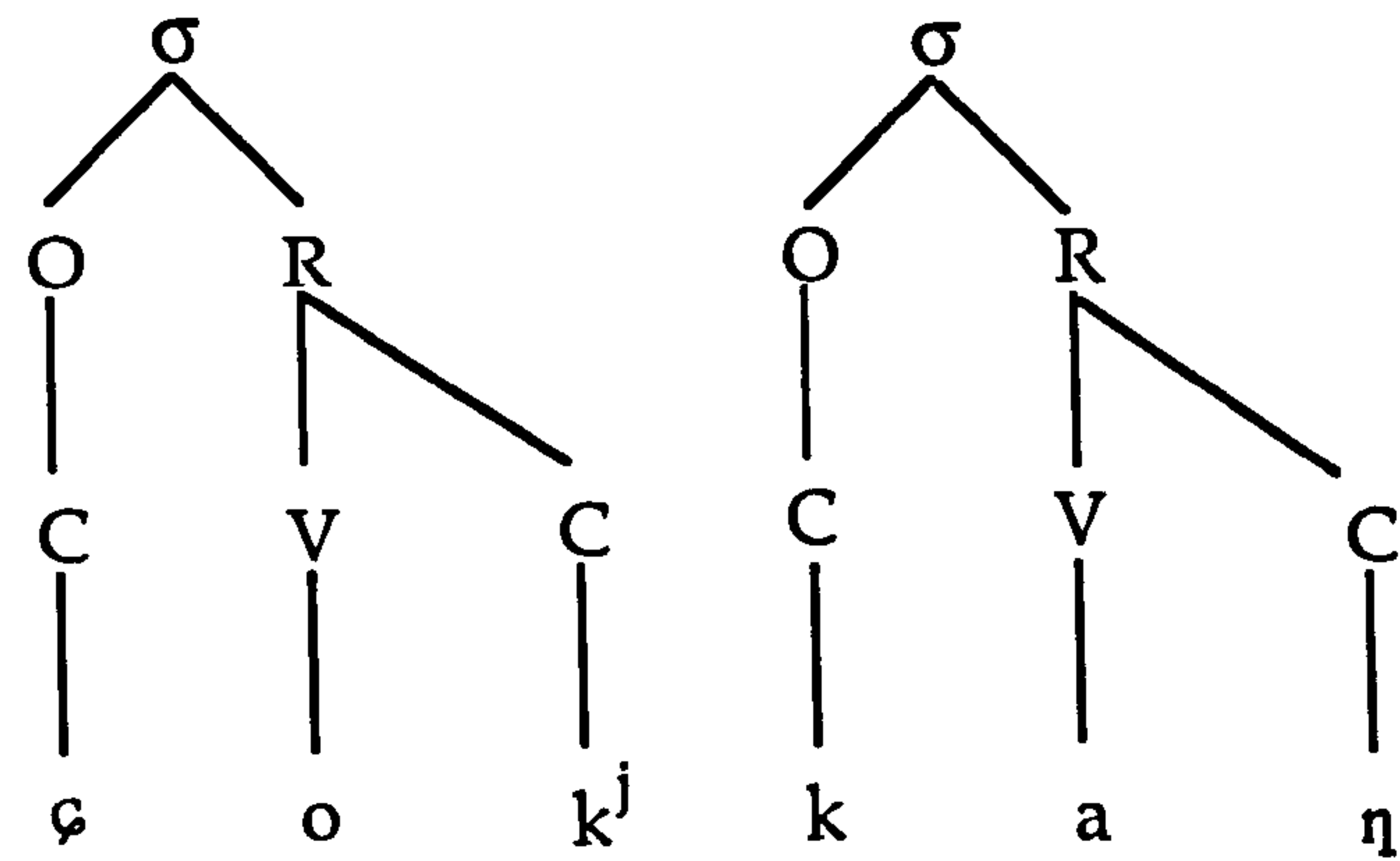
Similar processes to French schwa deletion may be applied to Japanese vowel devoicing. As described in Figure 3.1 in Chapter 3, the basic Japanese syllable structure is (C)(j)V(V)(C). Either a moraic nasal or voiceless obstruents which are the first part of consonant geminates can occupy the last optional C (the coda position). If we assume that the Japanese syllable structure can be extended to accept any kind of voiceless obstruent in addition to an initial part of geminate consonants in the coda position, vowel devoicing patterns might be able to be explained in terms of the syllable structure. When a vowel becomes voiceless and

loses its sonority, the preceding consonant becomes a floating segment and must be syllabified with the preceding syllable. For example, words with one devoiceable vowel *shokikan* /shoki-kaN/ [çokɨ-kaŋ] 'cabinet secretary' (7a) may be syllabified as (7b), similarly *tetsukabuto* /tetu-kabuto/ [tetsɯ-kabɯto] 'steel helmet' (8a) may be syllabified as (8b) ("-" indicates an internal word boundary).

(7a)

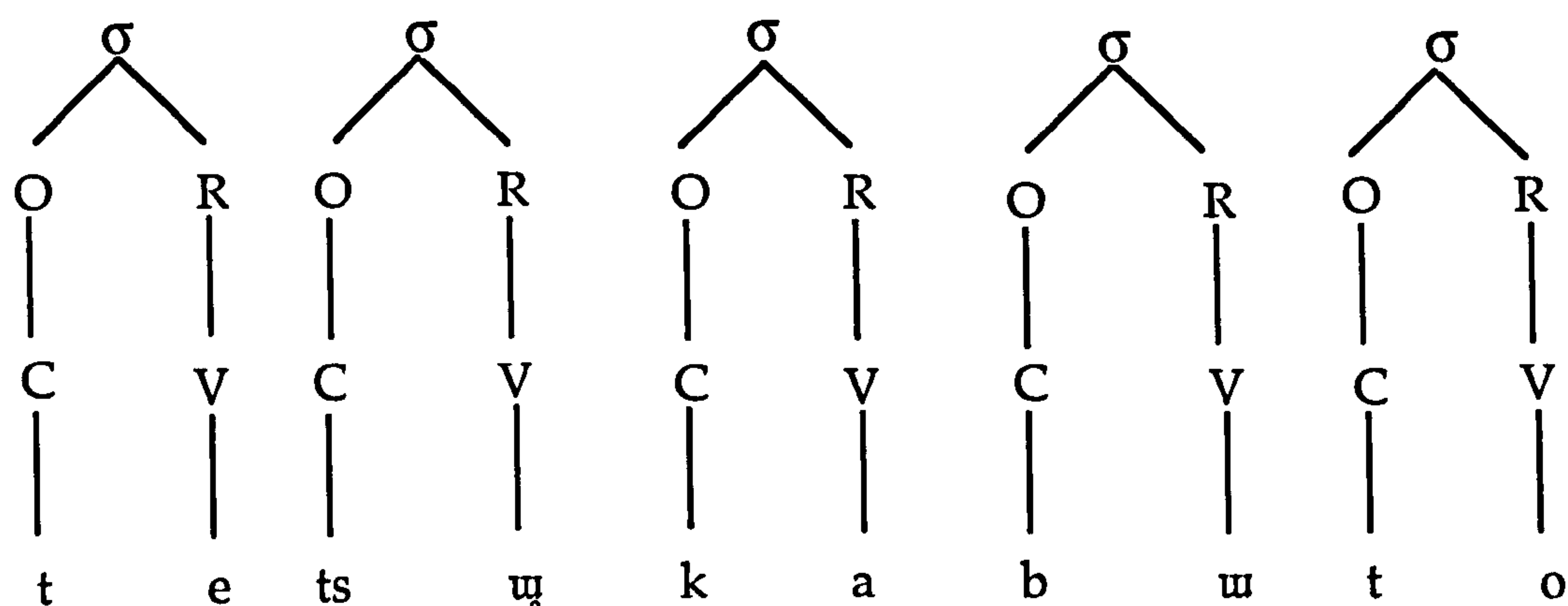


(7b)

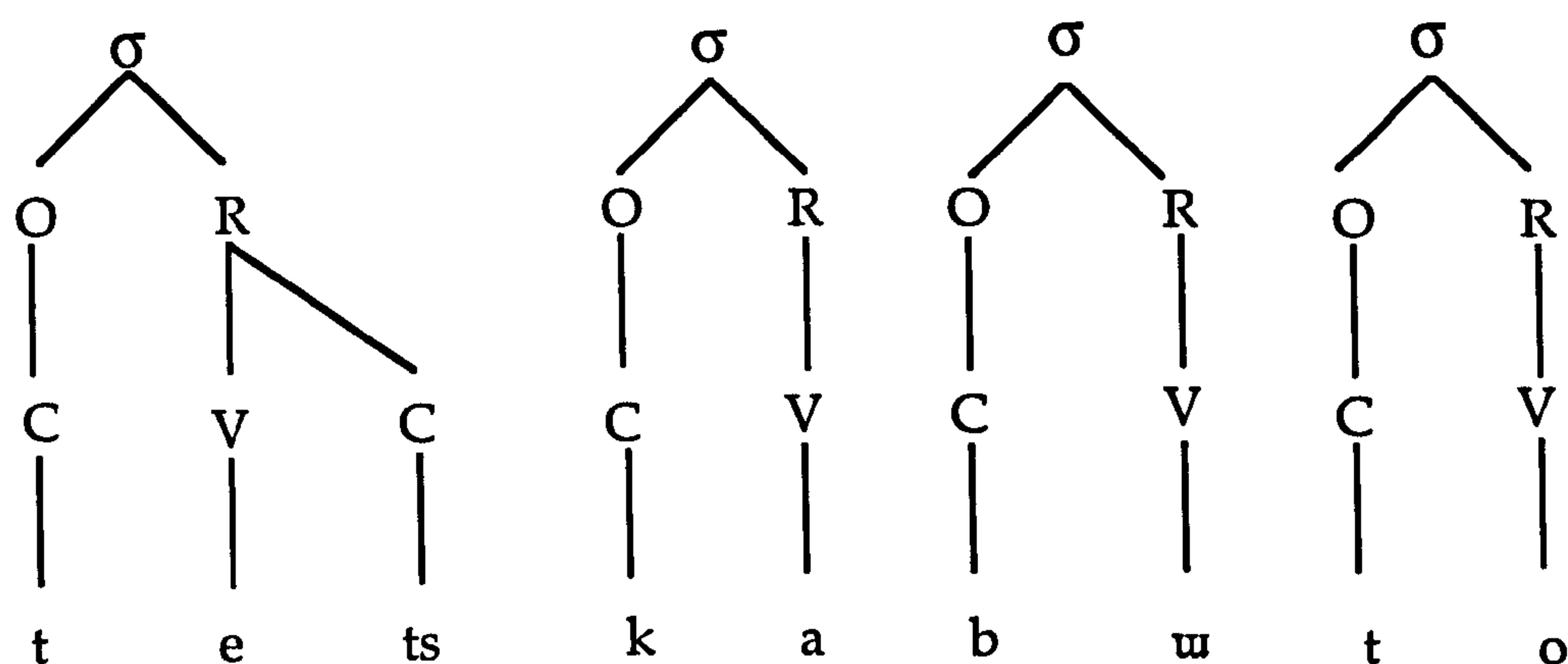


<sup>7</sup> [ŋ] is a moraic nasal.

(8a)



(8b)



When this syllabification is possible, a high vowel can undergo a weakening process and become voiceless, and ultimately may be deleted in casual speech or very fast speech.

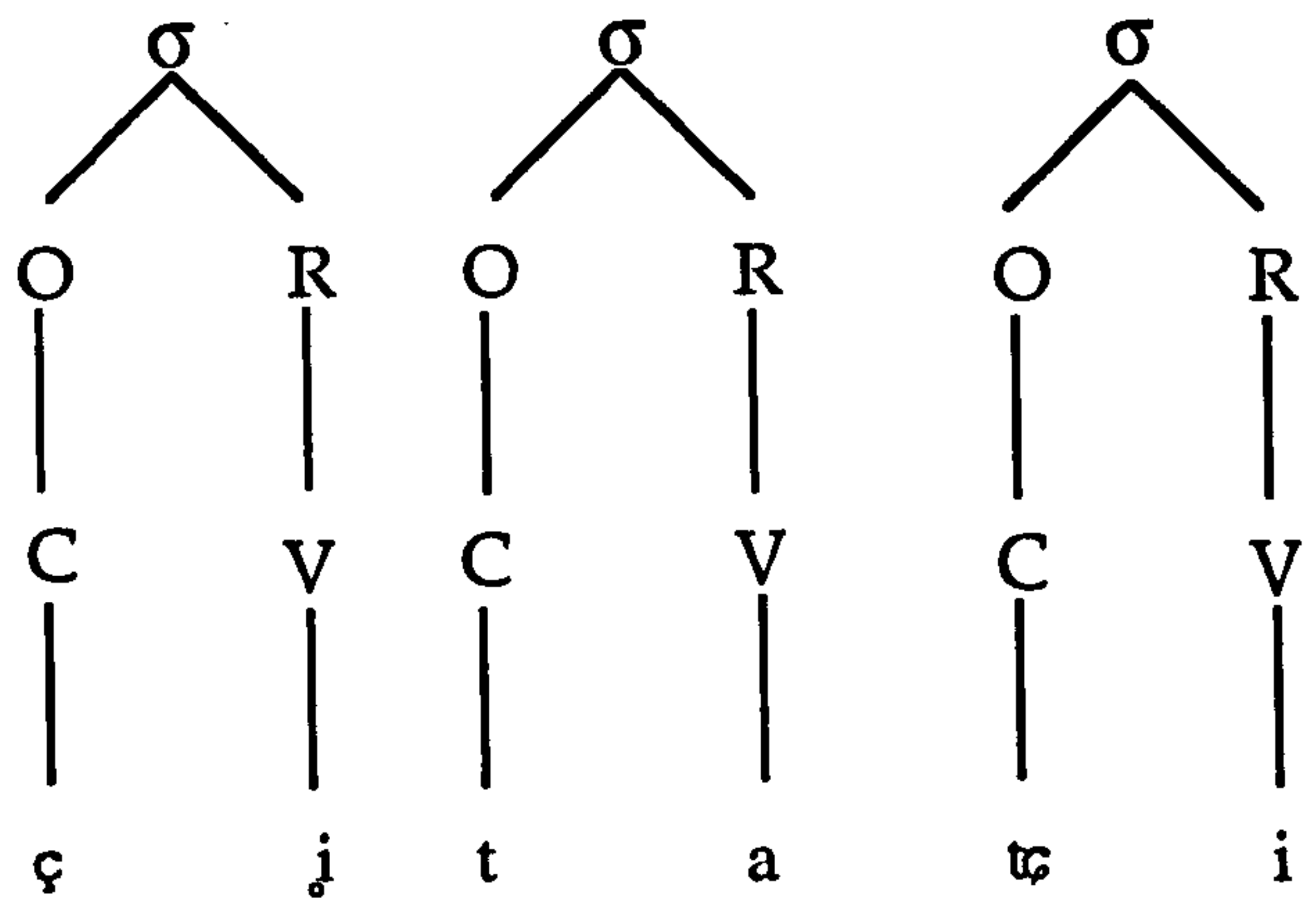
In word initial position, a similar rule to French 'onset accretion across schwa' might be applied to allow various sorts of voiceless obstruent clusters in the syllable onset position. When a vowel in the word initial syllable is devoiced and loses its sonority as in (9a), the preceding consonant is syllabified to the onset position of the following syllable as shown in (9b)<sup>8</sup>. However, this rule must be restricted to the

<sup>8</sup> The last vowel /i/ can also be devoiced. In this case, the preceding consonant [ɰ] is resyllabified to the coda position of the preceding syllable as shown in 9(c) in the footnote of the next page.

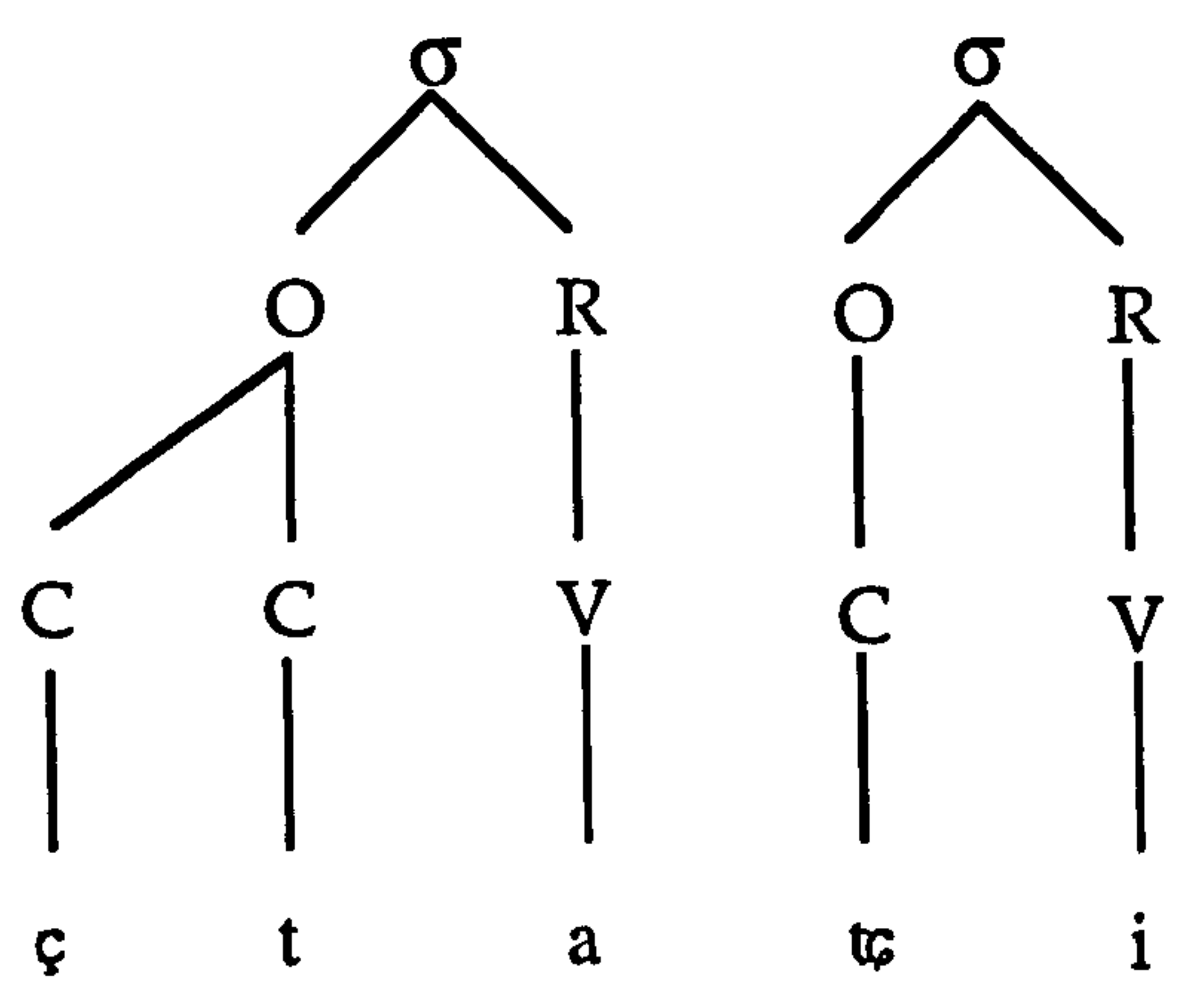


word initial position, including after an internal word boundary in cases of compound words.

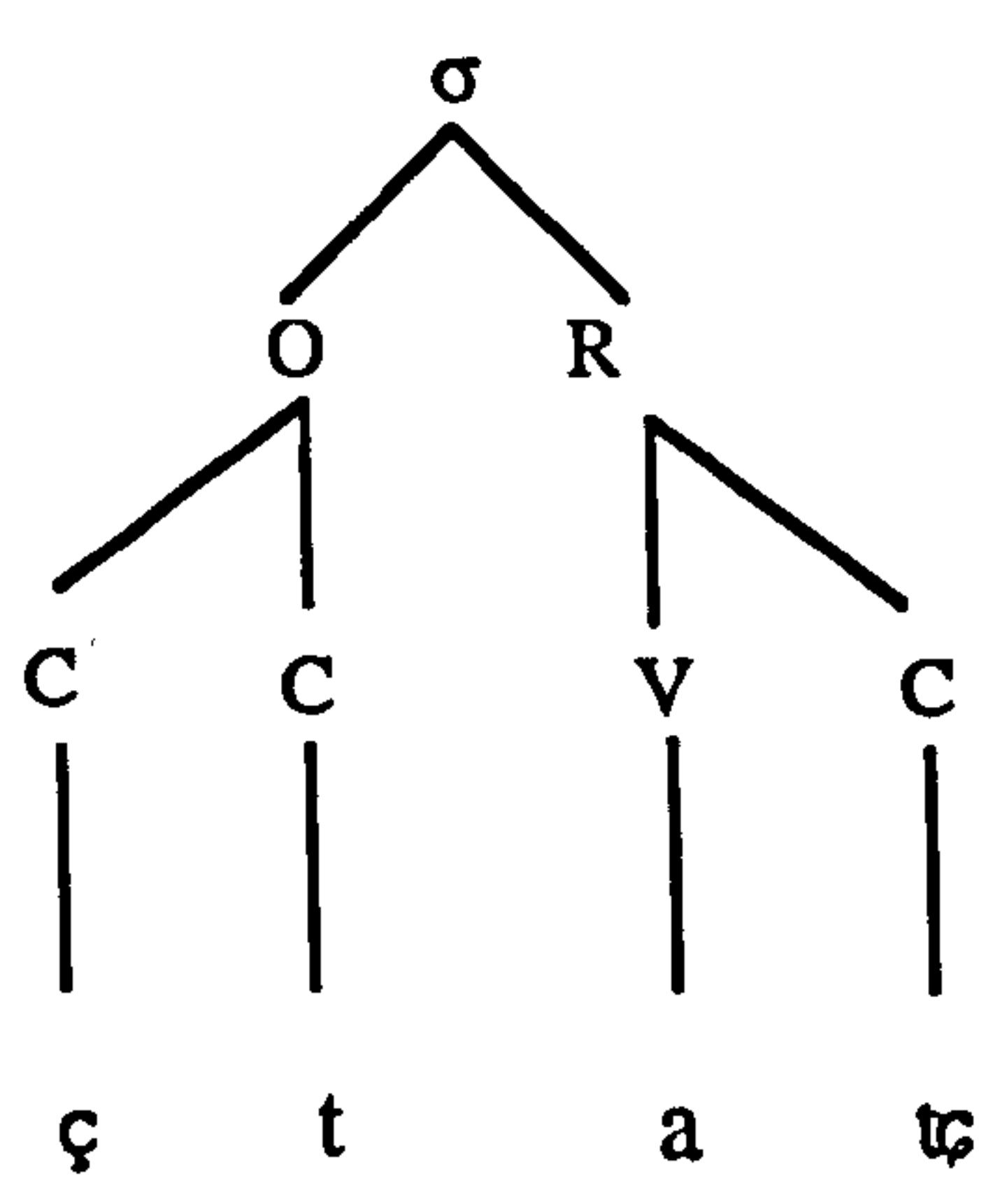
(9a) *Hitachi* /hitati/ [çitatɕi] 'place name'



(9b)



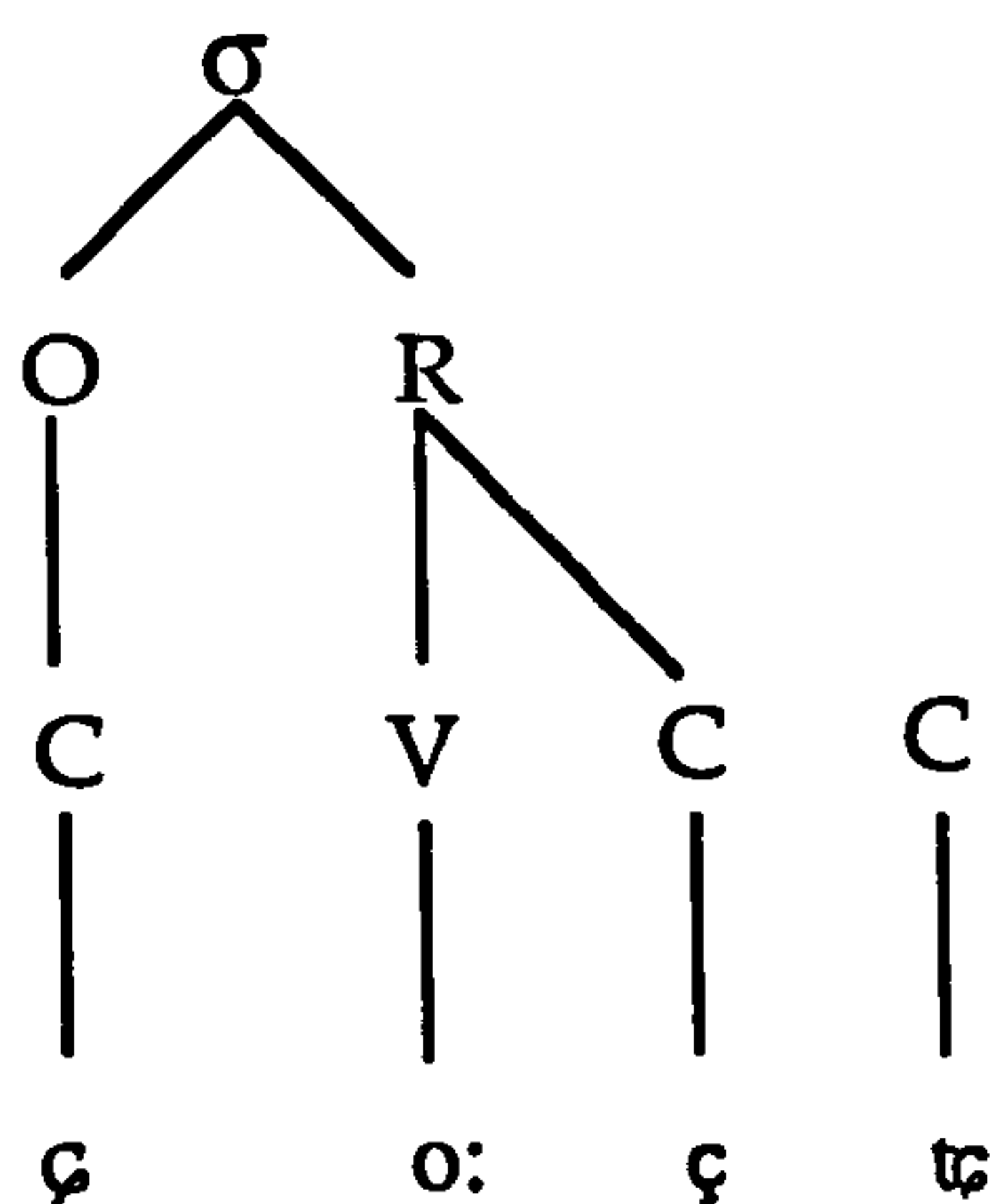
<sup>8</sup> (9c) [çitatɕi]



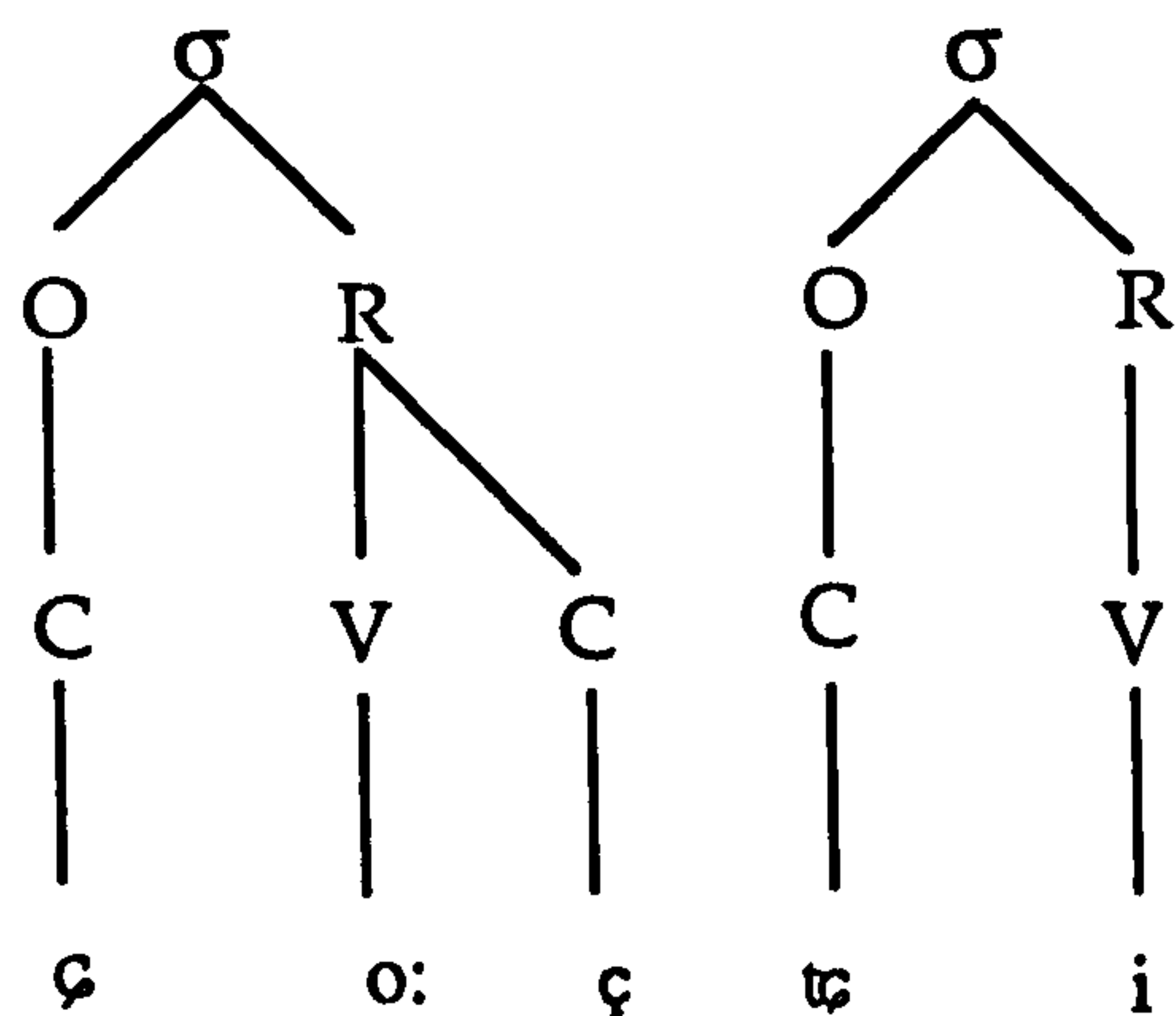
Therefore, vowel devoicing in the single devoicing environment is almost always possible as the preceding voiceless consonant can be syllabified either to the preceding syllable, or to the following syllable in the case of word initial position.

However, in the case of the consecutive devoicing environment, if all devoiceable vowels become voiceless and lose their sonority, not all preceding consonants can be syllabified to their adjacent syllables. For instance, underlined high vowels in *shoohiti* /çoohi-ti/ are devoiceable. However, in the case of *shoohiti* /çoohi-ti/ 'consumer belt', the pronunciation \*[ço:çj-tçj] with consecutive devoicing is not favoured in normal speech, because it is impossible to syllabify both consonants [ç] and [tç] preceding the voiceless vowels and satisfy Japanese syllable structure as shown in (10a). Instead, [ço:çj-tçi] with only one devoiced vowel as in (10b) is favoured.

(10a) \*[ço:çj-tçj]

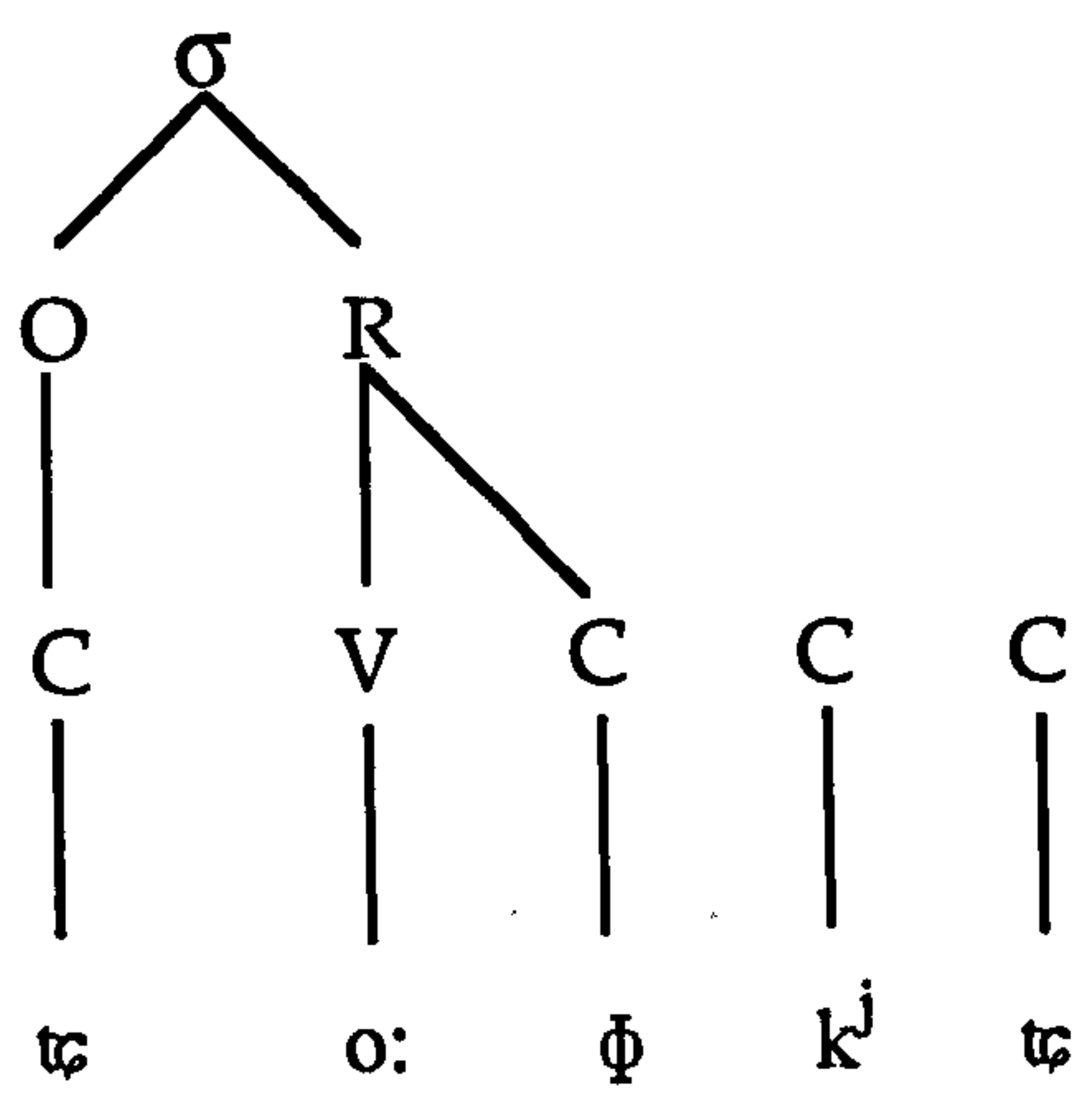


(10b) [ço:çj-tçi]

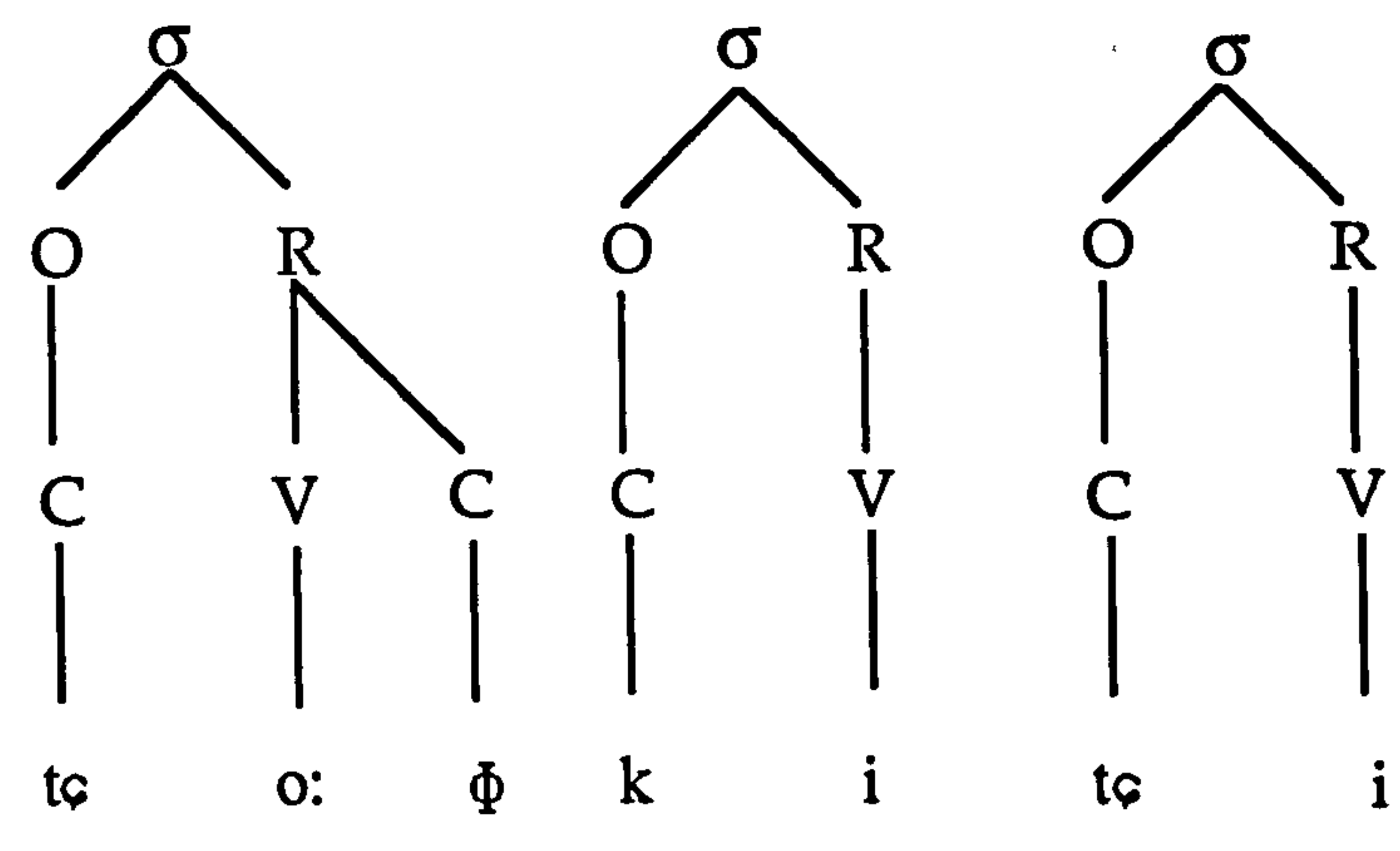


Similarly, in the case of *Choofukichi* /tjoohu-kiti/ 'the Chofu Base', a pronunciation \*[tço:çwç-kj-tçj] with three consecutive devoicings is not favoured because of unsyllabified consonants as in (11a), but pronunciations such as [tço:çwç-kitçi] and [tço:çwç-kjtçi] are possible as shown in (11b) and (11c) respectively. As for (11c), syllabification of [k<sup>j</sup>] to the onset position of the following syllable is allowed, since *Choofukichi* is a compound word (*Choofu* 'place name' + *kichi* 'base') and there is an internal word boundary.

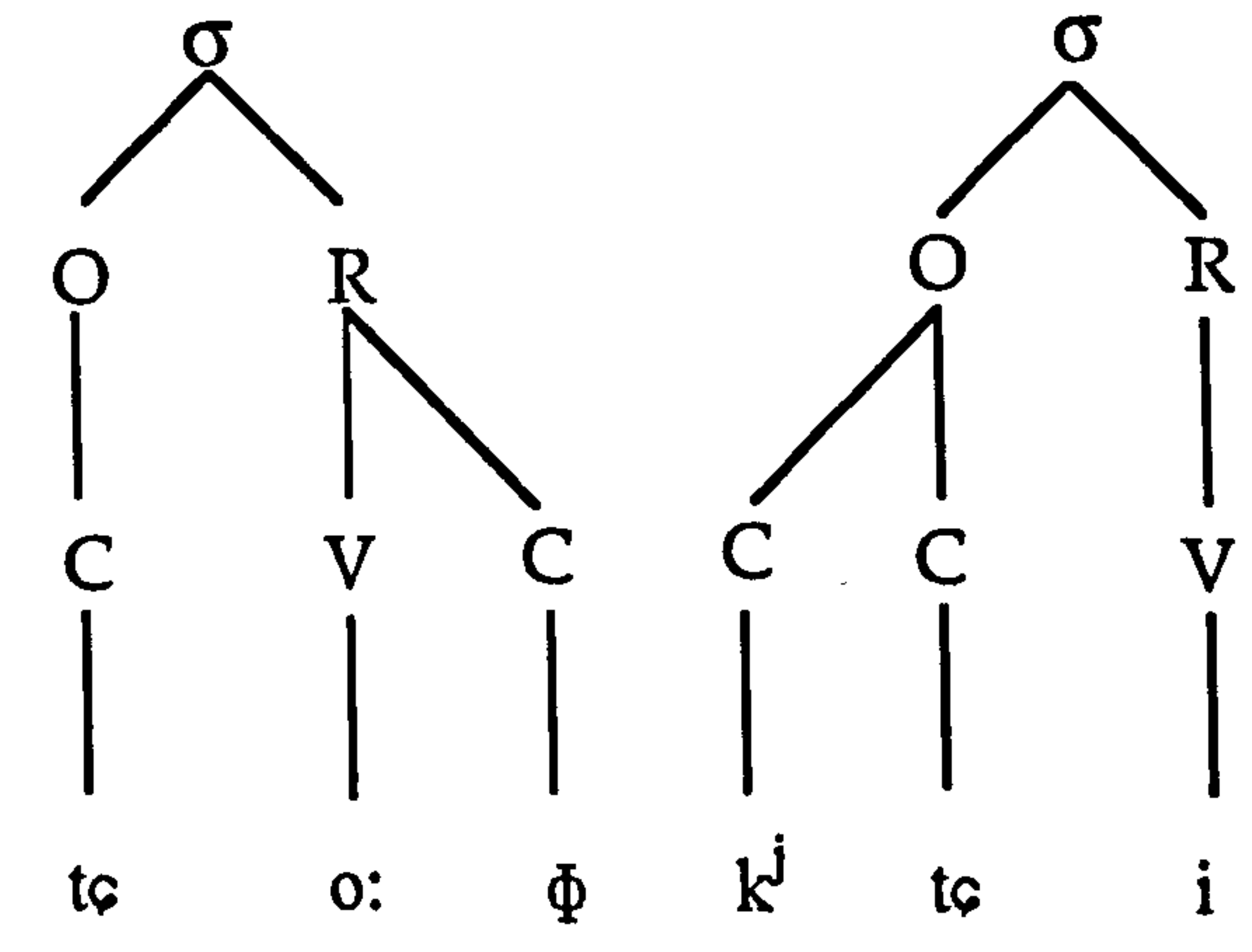
(11a) \*[tɕo:ɸɯ-kʲitɕi]



(11b) [tɕo:ɸɯ-kitɕi]



(11c) [tɕo:ɸɯ-kʲitɕi]





As demonstrated in Chapter 5, not all devoiceable vowels become voiceless in the consecutive devoicing environment in normal speech. The above syllabification rule works on most common voicing/devoicing patterns of each test word with consecutive devoicing sites listed in Table 5.14a - 5.14c in Chapter 5. There are some examples in which the syllabification of consonants is not possible, such as [fɯkɯ-to'çiŋ] 'the hub of the city', [kɯtsɯ-ko':dzo:] 'shoe factory' and [kasɯpj'-kai] 'the Caspian Sea', but they were not the most common pronunciations.

In the consecutive devoicing environment, the blocking factors of vowel devoicing proposed in Chapter 5, such as accentedness, become effective and subsequently an accented vowel may retain its sonority and syllabicity. Other devoiceable vowels may undergo the weakening process and eventually are devoiced only if syllabification of the preceding consonant is possible. Another crucial factor for devoicing may be the type of preceding consonant. As discussed and demonstrated in Chapters 5 and 6, preceding fricatives trigger devoicing more than preceding plosives. Therefore, when one of devoiceable vowels is accented and another devoiceable vowel in an adjacent syllable is preceded by a fricative, the devoiceable vowel following the fricative may be devoiced if the preceding fricative can be syllabified to its adjacent syllable. Actual voicing/devoicing patterns in the consecutive devoicing environment may be determined by the combination of various phonetic and phonological factors such as the syllabification rule, the type of preceding consonants, accentedness and position in an utterance.

Then, why is the syllable structure more important than the mora in vowel devoicing? As examined in Chapter 7, the mora seems to have some kind of phonetic reality in Japanese, although each mora does not have equal duration; the duration of a word is more or less determined by the number of morae in a word. The number of devoiced vowels did not have a significant effect on the duration of the entire word. Therefore, it might be perfectly acceptable to analyse vowel devoicing in terms of mora-timing, and retain the structure of the mora and devoiced vowels as /CV̥/.

Kubozono (1995) proposed that there is a tendency for Japanese to establish syllables of a bimoraic size. He suggested that the bimoraic foot is the most unmarked foot structure in the phonological theory of stress and rhythm, and the Japanese mora, whose duration is about 150 ms in normal speech, is too short to form a rhythmic unit, compared with inter-stress intervals in other languages, which fall within the range of 300-600 ms. Presenting examples from children's speech, adult's speech, historical phonology and foreign loan words, Kubozono showed that (i) phonologically heavy syllables (CVV and CVC) are dominant over light syllables (CV) in children's speech, (ii) there is lengthening of monomoraic syllables /CV/ to /CV:/ which is equivalent to /CVV/, (iii) in old Japanese, the CV syllables were phonetically much longer than the CV syllables in modern Japanese, and (iv) bimoraic heavy syllables, such as CVC and CVV, followed by a light syllable CV is commonly found in loan words. He proposed that all of these formations were operated in order to achieve the phonetic requirement on the duration of the syllable.

As discussed in Chapter 3, a typical Japanese syllable structure is (C<sub>1</sub>)(j)V<sub>1</sub>(V<sub>2</sub>)(C<sub>2</sub>)(C<sub>3</sub>). /C<sub>2</sub>/ can be either a voiceless consonant, only where it is the first element of geminate consonants, or it can be a moraic nasal. /C<sub>3</sub>/ is possible only when /C<sub>2</sub>/ is a moraic nasal and the /C<sub>3</sub>/ is the first element of geminate consonants as in /roNdoNkko/ 'Londoner' and /hoNtte/ '(someone says that) books are ...'. The only acceptable consonant clusters in the syllable initial position are /Cj/ sequence. If Japanese actually allows other clusters of voiceless consonants in syllable initial position and /C<sub>2</sub>/ can be any voiceless consonant, the vowel devoicing processes can be explained in terms of the new more flexible syllable structure. When a vowel was devoiced and lost its sonority, the consonant preceding the devoiced vowel was resyllabified to the coda position of the preceding syllable. Word initially, when a vowel was devoiced, the preceding consonant was resyllabified to the onset position of the following syllable occupying the /C<sub>1</sub>/ position. A vowel underwent the devoicing process only when these resyllabifications were possible.

Kubozono's proposal and the results of devoicing patterns in the consecutive devoicing environment suggest that the syllable as well as the mora plays an important role in Japanese phonology. The results in Chapter 6 showed that the mora is, without a doubt, a rhythmic unit in Japanese. However, the experimental results in this chapter indicated that vowel devoicing processes cannot be worked out solely from their phonetic environments. The above phonological analysis of Japanese vowel devoicing is by no means perfect, but suggests that vowel devoicing in Japanese may need to be examined in relation to Japanese syllable structure.

## 8.6 Conclusions

High vowels between voiceless consonants are not always devoiced sometimes remaining voiced in this typical devoicing environment. However, these voiced vowels are acoustically different from vowels in non-devoicing environments. Earlier experimental results in Chapters 5, 6 and 7 showed that voiced high vowels in a typical devoicing environment were often not actually fully voiced and were reduced in duration. The experimental results in this chapter also demonstrated that these voiced devoiceable vowels were not only shorter in duration but also weaker in intensity. In other words, it seemed to be more natural for high vowels to undergo the devoicing process between voiceless sounds; so when they did remain voiced, they were acoustically different from ordinary vowels in the non-devoicing environment.

As for the cause of the processes, the Japanese vowel devoicing process may be the result of laryngeal assimilation of high vowels and their adjacent voiceless consonants similar to the vowel reduction processes in other languages as proposed by Jun and Beckman (1993 and 1994).

It is possible for vowels in typical devoicing environments to remain voiced, but in these cases they are often weakened. Vowel devoicing is not a categorical rule between voiced and voiceless, with the phonetic realisation of devoiceable vowels



varying from fully voiced to completely voiceless. This suggests that vowel devoicing is part of a vowel weakening process and the final state of the process is completely voiceless vowels. When there is no blocking factor present, it is normal for the vowel to undergo fully the weakening process resulting in complete devoicing. This process was encouraged when the speaking rate increased. However, the weakening process is constrained by other factors such as devoicing conditions, meaning that sometimes weakening fails to occur, which results in the voicing of a devoiceable vowel. However, though the vowel may be voiced, its intensity is weaker than that of ordinary vowels.

The data from consecutive devoicing cases suggested that the vowel weakening process may be influenced by Japanese syllable structure. Two different mechanisms, namely phonetic and phonological processes, seem to control Japanese vowel devoicing depending on the environment. When vowel devoicing is not favoured because of the constraint on the syllable structure, vowel weakening does not occur, and the intensity of a voiced devoiceable vowel remains as high as that of non-devoiceable vowels.

It should also be noted that vowel weakening in Japanese primarily affects vowel intensity, then duration, and the quality of the vowels remained relatively unchanged regardless of the intensity level of the vowel. The vowel weakening process in Japanese seems to work differently from the similar process in stress-timed languages.

# Chapter 9

## Conclusions

In this study, Japanese vowel devoicing was examined in relation to phonetic and phonological factors. Experiments investigated which factors triggered devoicing and which blocked devoicing. Examining devoicing in combination with different factors helped to understand more fully the mechanisms of Japanese vowel devoicing. Close observation of vowels in devoicing and non-devoicing environments discovered the gradient nature of devoicing, and the intensity measurement of voiced devoiceable vowels demonstrated that Japanese vowel devoicing is actually part of the vowel weakening process. A syllable-based analysis of devoicing lead to a better understanding of the weakening process and clarified patterns of devoicing in the consecutive devoicing environment. The experiments also investigated effects of various phonetic and phonological factors which affect Japanese vowel devoicing processes, and showed that devoicing is the result of a complex interaction of these factors.

The study investigated Japanese vowel devoicing in two separate conditions, namely single and consecutive devoicing environments, because it is only in the consecutive environment that some devoiceable vowels consistently fail to undergo any devoicing process. The effect of blocking factors for devoicing, such as accent on a vowel and the presence of a following internal word boundary, were not significant in the single environment, where vowel devoicing was almost compulsory. However, in the consecutive environment, these blocking factors were effective and prevented some devoiceable vowels from becoming voiceless (Chapter 5). The type of blocking

factors, and whether there was one or more than one factor present did not alter their overall effectiveness. The type of preceding consonant and speaking tempo also affected devoicing in the consecutive environment; preceding fricatives triggered devoicing significantly more than preceding plosives and affricates (Chapter 5), while devoicing occurred more at faster tempi than slower tempi (Chapter 6).

The type of preceding consonant, speaking rate and the devoicing condition were also found to be significant when studying the intensity of vowels in the devoicing environment. Intensity measurements of voiced vowels in devoicing and non-devoicing environments showed that intensity weakening of voiced devoiceable vowels occurred in the single devoicing environment, but not in the consecutive devoicing environment (Chapter 8). In the single environment, the majority of devoiceable vowels were completely devoiced, but when some devoiceable vowels remained voiced, their intensities were weakened greatly compared with those of non-devoiceable vowels. The intensities of voiced devoiceable vowels in the single environment were weakened more at faster tempi than at slower tempi. In the consecutive environment, however, there was no significant weakening of intensity of voiced devoiceable vowels even when the speaking tempo increased. The results imply that it is more natural for devoiceable vowels in the single environment to be devoiced. Therefore, when they remain voiced, they are acoustically different from vowels in non-devoicing environments.

The intensity reduction of voiced deviceable vowels occurred only in the single devoicing environment, where vowel devoicing was favoured. In the same environment, vowel intensity was generally lowered more at faster tempi, at which tempi devoicing rates were higher. Since vowel devoicing and intensity reduction occurred in the same environment, this suggests that vowel devoicing is part of the vowel weakening process and the ultimate stage of the process is a completely voiceless vowel. Vowels in the devoicing environment can be realised at any stage of the weakening process. This explains the optionality of vowel devoicing in the typical devoicing environment.



The weakening process of Japanese high vowels occurs in similar environments to those of the vowel reduction processes in other languages, as discussed in Chapter 2. Vowel devoicing occurs in many languages as part of the vowel reduction process, where a vowel is at first centralised and reduced in duration in an unstressed position, often in fast or casual speech. The Japanese vowel weakening process does not, however, involve centralisation of vowel quality (Chapter 8). It involves a reduction of duration and intensity, but the reduced duration of devoiced morae were adjusted for at the word level, so that the duration of the entire word was still determined only by the number of morae in a word remaining unchanged regardless of whether a vowel is voiced or voiceless (Chapter 7). The experimental results confirmed the complex mechanism of durational adjustment of Japanese words, and the importance of the mora in speech rhythm in Japanese as discussed in Chapter 3.

However, the syllable may play a significant role in the whole vowel weakening process. The difference between single and consecutive devoicing environments is that when a vowel is devoiced and loses its sonority, the single environment phonetically creates only one sequence of voiceless consonants, whereas the consecutive environment can potentially create sequences of more than one voiceless consonant. A possible constraint which prevents a sequence of more than one consonant is syllable structure.

In most languages the most common environment for vowel devoicing is a high vowel adjacent to a voiceless consonant (Chapter 2). This is a phonetically ideal environment, because a high vowel is inherently short in duration and the glottis is pulled apart during the production of a voiceless consonant. Japanese vowel devoicing also occurs in similar environments (Chapter 4). Thus, Japanese vowel devoicing is primarily a process of overlapping gestures of laryngeal muscles, between a short vowel and its adjacent voiceless sounds. However, the laryngeal assimilation does not operate for all devoiceable vowels in the consecutive environment. If the gestural overlap of the laryngeal muscles is the only cause of devoicing, in theory all devoiceable vowels should become voiceless.

If a new syllable structure by vowel devoicing is permitted, a devoiceable vowel undergoes the devoicing process. Therefore, devoiceable vowels in the single environment, where resyllabification of the preceding consonant is always possible, are almost always devoiced. However, in the consecutive environment, devoicing all potentially devoiceable vowels can violate the constraints on syllable structure, and therefore it is only devoiceable vowels which are in the more ideal conditions phonetically which become voiceless, while the other devoiceable vowels remain voiced. When the vowels are in phonetically equal conditions, the blocking factors become effective and block some devoiceable vowels from becoming voiceless, so as not to violate the syllable structure (Chapter 4).

As long as the syllable structure permitted it, preceding fricatives triggered devoicing of the following vowel significantly more than the preceding plosives or affricates in the consecutive devoicing environment. There are physiological studies which confirm the positive effect of preceding fricatives on vowel devoicing. Preceding fricatives provide physiologically more ideal conditions for a vowel to be devoiced than preceding plosives (Kuehn and Moll, 1976, Weitzman et al, 1976, Yoshioka et al, 1982, and Simada et al, 1991) (Chapter 6).

As demonstrated in Chapter 5, with vowel devoicing there is not a clear cut distinction between voiced or voiceless, but the process showed many cases of partially voiced vowels. Vowel devoicing is also not categorical. Devoiceable vowels sometimes fail to become voiceless. Neither traditional phonological account of devoicing based on the categorical process (Kawakami, 1976, Vance, 1987, Maekawa, 1989, etc.), nor phonetic accounts based on the gestural overlap analysis (Jun, 1990, and Jun and Beckman, 1993 and 1994) satisfactorily explain the whole process of vowel devoicing in Japanese. The devoicing environments provide phonetically ideal conditions for the vowel weakening process, and a vowel eventually become voiceless as part of the process. However, various blocking factors sometimes prevent the vowel from becoming voiceless. The overall vowel weakening process appears to be constrained by Japanese syllable structure. This implies the importance of syllable in Japanese

phonology as suggested by Shibatani (1987) and Kubozono (1995) (Chapters 4 and 8).

When a high vowel is surrounded by voiceless sounds, the vowel laryngeal gestures naturally overlap with those of the voiceless consonants, and results in the devoicing of the vowel. Phonetic factors such as a preceding fricative and faster speaking rate also promote devoicing. However, the devoicing process must satisfy phonological constraints such as the syllable structure. Where the syllable structure permits, devoiced vowels become voiceless. Even factors which are supposed to block devoicing did not prevent the vowel from undergoing the weakening process and in most cases the vowel became voiceless. In the environments where consecutive devoicing violates the syllable structure, preceding fricatives trigger devoicing, and blocking factors become effective preventing some devoiced vowels from becoming voiceless. The syllable structure may be violated in fast speech, as a faster speaking tempo triggers vowel devoicing significantly more than a slow tempo, which creates many examples of devoicing in consecutive morae.

The devoicing processes may be worked out by ranking the triggering and blocking factors of devoicing as suggested in Optimality Theory (Prince and Smolensky, 1993). Ranking the factors may provide patterns of voicing and devoicing in the consecutive environment. However, in addition to the blocking factors studied in this experiment, there are other important factors such as word frequency and psychological heaviness (see Chapter 2). These other factors may have had some influence on the results in this study and additional research is needed to assess the exact effects of such factors to better understand the mechanisms of Japanese vowel devoicing.



## REFERENCES

- Abe, Y. (1986) Metrical structure and compounds in Japanese, pp. 5-52, in Imai, I. and Saito, M. (Eds); *Issues in Japanese Linguistics*, Foris Publications.
- Abercrombie, D. (1967) *Elements of General Phonetics*, Edinburgh University Press, Edinburgh.
- Armogost, J. L. (1986) Three exceptions to vowel devoicing in Commanche, *Anthropological Linguistics* 28 (3) : 255-65.
- Árnason, K. (1986) The segmental and suprasegmental status of preaspiration in modern Icelandic, *Nordic Journal of Linguistics*, Vol. 9, No.1: pp. 1-23.
- Beckman, M. E. (1982) Segment duration and the 'Mora' in Japanese, *Phonetica* 39 : 113-135.
- Beckman, M. E. (1993) When is a syllable not a syllable?, paper presented at the International Forum of Speech Recognition and Phonology, Dokkyo University, Saitama, Japan, 18-19 December, 1993.
- Beckman, M. E. and Shoji, A. (1984) Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese, *Phonetica* 4:61-71.
- Bell, A. (1978) Syllabic consonants, in Greenberg, J. H. (Eds.) *Universals of human language Vol. 2 Phonology*, Stanford University Press, Stanford, California.
- Bloch, B. (1970) *Bernard Bloch on Japanese*, (Ed.) by Miller, A. Yale University Press, New Haven.
- Borden, G. J. and Harris, K. S. (1984) *Speech Science Primer, 2nd Ed.*, Williams & Wilkins, Baltimore, Maryland.
- Browman, C. P. and Goldstein, L. (1989) Articulatory gestures as phonological units, *Phonology* 6: pp. 201-251.
- Browman, C. P. and Goldstein, L. (1990a) Tiers in articulatory phonology, with some implications for casual speech, pp. 341-376, in Kingston, J. and Beckman, M. E. (Eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, Cambridge University Press, Cambridge.
- Browman, C. P. and Goldstein, L. (1990b) Gestural specification using dynamically-defined articulatory structures, *Journal of Phonetics* 18: 299-320.
- Campbell, N. (1992) Segmental elasticity and timing in Japanese speech, pp. 403-418, in ATR Auditory and Visual Perception Research Laboratories (Ed.) *Speech Perception, Production and Linguistic Structure*, Ohmsha, Tokyo.

- Campbell, N. and Sagisaka, Y. (1991) Moraic and syllable-level effects on speech timing, *Journal of Electronic Information Communication Engineering*, SP 90-107:35-40.
- Canonge, E. D. (1957) Voiceless vowels in Comanche, *International Journal of American Linguistics*, 23(2): 63-67.
- Casagrande, J. B. (1954) Comanche Linguistic Acculturation : I, *International Journal of American Linguistics*, 20(2) : 140-151.
- Catford, J. C. (1977) *Fundamental Problems in Phonetics*, Edinburgh University Press.
- Chen, M. (1970) Vowel length variation as a function of the voicing of the consonant environment, *Phonetica*, 22: 129-159.
- Coleman, J. (1991) Non-pitch exponents of accent and structure in Japanese, *York Papers in Linguistics* 15: 41-92.
- Crystal, D. (1985) *A Dictionary of Linguistics and Phonetics* (2nd Ed.), Blackwell, Oxford.
- Dauer, R. M. (1980a) *Stress and Rhythm in Modern Greek*, Ph.D thesis, University of Edinburgh.
- Dauer, R. M. (1980b) The reductions of unstressed high vowels in modern Greek, *Journal of the International Phonetic Association*, 10(1-2): 17-27.
- Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (1952) An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic patterns, *Word*, 8: 195-210.
- Donegan, P. J. (1985) *On the Natural Phonology of Vowels*, Garland Publishing, New York.
- Fant, G. (1960) *Acoustic Theory of Speech Production*, 's-Gravenhage.
- Fry, D. B. (1979) *The Physics of Speech*, Cambridge University Press, Cambridge.
- Fukai, K. (1979) Nihon-go ni okeru boin no jak-ka, musei-ka, shoomezsu ni tsuite, *Risshoo University Bulletin of General Studies*, Risshoo University, 13: 45-55.
- Fujisaki, H. (1989) Nihon-go no onchou no bunseki to moderu-ka: Go-akusento, tougo kouzou, danwa kouzou to onchou tono kankei, pp. 266-297, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (I)*, Meiji Shoin, Tokyo.
- Fujisaki, H. and Sugito, M. (1977) Onsei no butsuri-teki seishitsu, pp. 63-106, in Ohno, S. and Shibata, T. (Eds.) *Iwanami Kooza Nihon-go: Vol. 2 On'in*, Iwanami Shoten, Tokyo.
- Fujisaki, H. and Higuchi, N. (1980) Temporal organization of segmental features in Japanese disyllables, *Journal of the Acoustical Society of Japan*, (E) Vol. 1, No. 1: 25-30.



Gay, T. (1978) Effect of speaking rate on vowel formant movements, *Journal of the Acoustical Society of America*, Vol. 63, No. 1, January: 223-230.

Gay, T. (1979) Coarticulation in some consonant-vowel and consonant cluster-vowel syllables, pp. 69-76, in Lindblom, B. and Ohman, S. (Eds.) *Frontiers of Speech Communication Research*, Academic Press, London.

Gay, T. and Hirose, H., (1973) Effect of speaking rate on labial consonant production: a combined electromyographic/high-speed motion picture study, *Phonetica* 27: 44-56.

Gay, T., Ushijima, T., Hirose, H. and Cooper, F. S. (1974) Effect of speaking rate on labial consonant-vowel articulation, *Journal of Phonetics* 2: 47-63.

Gimson, A. C. (1980) *An introduction to the Pronunciation of English*, 3rd Ed., Edward Arnold, London.

Gobl, C. and Ni Chasaide, A. (1988) The effects of adjacent voiced/voiceless consonants on the vowel voice source: a cross language study, *Speech Transmission Laboratory Quarterly Progress and Status Report*, pp. 23-59, Sweden.

Griffen, T. D. (1983) The Swabian voiceless vowels, *Word*, 34(3): 145-173.

Griffen, T. D. (1985) *Aspects of Dynamic Phonology*, John Benjamins Publishing Co., Amsterdam.

Han, M. S. (1962a) The feature of duration in Japanese, pp. 65-80, The Phonetic Society of Japan (Ed.) *The Study of Sounds* Vol. 10, Tokyo.

Han, M. S. (1962b) Unvoicing of vowels in Japanese, pp. 81-100, The Phonetic Society of Japan (Ed.) *The Study of Sounds* Vol. 10, Tokyo.

Han, M. S. (1994) Acoustic manifestations of mora timing in Japanese, *Journal of the Acoustical Society of America* 96 (1), July: 73-82.

Haraguchi, S. (1982) An autosegmental theory: Its expansion and extension, *Journal of Linguistic Research* Vol. 2, Part 4 : 59-80.

Hardcastle, W. J. (1976) *Physiology of Speech Production: An Introduction for Speech Scientists*, Academic Press, London.

Harris, K. S. (1979) Vowel duration change and its underlying physiological mechanisms, *Haskins Laboratories: Status Report on Speech Research*, SR-57.

Hasegawa, N. (1979) Casual speech vs. fast speech, pp. 126-137, in Clyne, P. R., Hanks, W. F. and Hofbauer, C. L. (Eds.) *Papers from the Fifteenth Regional Meeting Chicago Linguistic Society*, 19-20 April, 1979, Chicago Linguistic Society, University of Chicago, Chicago.

Hashimoto, M. (1977) On'in no taikai to koozoo, pp. 1-29, Ohno, S. and Shibata, T. (Eds.) *Iwanami Kooza Nihon-go: vol. 2 On'in*, Iwanami Shoten, Tokyo.

Hatch, E. and Farhady, H. (1982) *Research Design and Statistics for Applied Linguistics*, Newbury House Publishers, Cambridge, Massachusetts.



Hattori, N. (1989) *Mechanisms of Word Accent Change : Innovations in Standard Japanese*, Ph.D thesis, University College London.

Hattori, S. (1955) On'in-ron, reprinted in Shibata, T., Kitamura, H. and Kindaichi, H. (Eds.) (1980) *Nihon no Gengo-gaku: Vol. 2 On' in*, pp.157-185, Taishuukan Shoten Tokyo.

Hattori, S., Yamamoto, K., Kohashi, Y. and Fujimura, O. (1957) Nihon-go no boin, reprinted in Shibata, T., Kitamura, H. and Kindaichi, H. (Eds.) (1980) *Nihon no Gengo-gaku: Vol. 2 On' in*, pp. 68-99, Taishuukan Shoten, Tokyo.

Higuchi, N., Fujisaki, H. and Sekimoto, S. (1978) Production and perception of segmental durations in spoken Japanese, *Journal of the Acoustical Society of America*, Vol. 64: S113.

Hirayama, T. (1985) Zen-Nihon no hatsuon to accent, pp.37-69, in Appendix to NHK (Ed.), *NHK Accent Dictionary*, Tokyo.

Hirose, H. (1989) Hatsuon no seiriteki shikumi, pp. 64-84, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (I)*, Meiji Shoin, Tokyo.

Hoequist Jr, C. (1983) Syllable duration in stress-, syllable- and mora-timed languages, *Phonetica* 40 : 203-237.

Hooper, J., B. (1972) A note on inserted and deleted vowels, *Working Papers on Language Universals*, 10: 141-144.

Hyman, L. M. (1970) The role of borrowing in the justification of phonological grammars, *Studies in African Linguistics* 1: 1-48.

Hyman, L. M. (1975) *Phonology: Theory and Analysis*, Holt, Rinehart and Winston, New York.

Imai, K. (1980) Onsei-gaku teki hikaku, pp. 7-67, in Kunihiro, T (Ed.) *Nichi-Eigo Hikaku Kooza*, Taishuukan Shoten, Tokyo.

The International Phonetic Association (1984) *The Principles of the International Phonetic Association*, London.

Johns, C., M. (1972) *Slips of the Tongue in Portuguese*, M.Litt Dissertation, University of Edinburgh.

Jōo, H. (1977) Gendai Nihon-go no on'in, pp. 107-145, in Ohno, S. and Shibata, T. (Eds.) *Iwanami Kooza Nihon-go: Vol. 2 On'in*, Iwanami Shoten, Tokyo.

Jōo, H. (1989) Onsei kigou, pp. 21-40, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (I)*, Meiji Shoin, Tokyo.

Jun, S.-A. (1993) *The Phonetics and Phonology of Korean*, Ph.D dissertation, Department of Linguistics, The Ohio State University.

Jun, S.-A. and Beckman, M. E. (1993) A gestural-overlap analysis of vowel devoicing in Japanese and Korean, Paper presented at the 1993 Annual Meeting of the Linguistic Society of America, los Angeles, California, 7-10 January, 1993.

- Jun, S.-A. and Beckman, M. E. (1994) Distribution of devoiced high vowels in Korean, *Proceedings of the 1994 International Conference on Spoken Language Processing*, Vol. 2: pp. 479-482, Yokohama.
- Kagaya, R. (1971) Laryngeal gestures in Korean stop consonants, *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 5: 15-23, Tokyo University.
- Kagaya, R. (1974) A fiberscopic and acoustic study of the Korean stops, affricates and fricatives, *Journal of Phonetics* 2: 161-181.
- Kaiki, N., Takeda, K. and Sagisaka, Y. (1990) Statistical analysis for segmental duration rules in Japanese speech synthesis, *Proceeding of the 1990 International Conference on Spoken Language Processing*, pp. 17-20, Kobe.
- Kaiki, N. and Sagisaka, Y. (1992) Analysis of pause duration based on local phrase structure (in Japanese), *Journal of Electronic Information Communication Engineering* SP91-130 : 63-69.
- Kakita, K. (1977) Hatsuon no kikoo, pp. 31-61, Ohno, S. and Shibata, T. (Eds.) *Iwanami Kooza Nihon-go: Vol. 2 On'in*, Iwanami Shoten, Tokyo.
- Katada, F. (1990) On the representation of Moras: evidence from a language game, *Linguistics Inquiry*, pp. 641-646.
- Kato, H., Tsuzaki, M. and Sagisaka, Y. (1992) Acceptability and discrimination threshold for distortion of segmental duration in Japanese words, *Proceedings of International Conference on Spoken Language Processing* (Banff).
- Kawakami, S. (1977) *Nihon-go Onsei Gaisetsu*, Ohfuusha, Tokyo.
- Kawasaki, H. (1983) Onsei no jikan seigyō ni kansuru moderu to jissoku deeta: Nihon-go to Ei-go ni okeru isochrony ni tsuite, *Journal of the Acoustic Society of Japan*, Vol. 39, No. 6: 389-397.
- Keating, P. and Huffman, K. (1984) Vowel variation in Japanese, *Phonetica* 41 : 191-207.
- Keating, P. (1985) Universal Phonetics and the organization of grammars, pp. 115-132, in Fromkin, V. A. (Ed.) *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, Academic Press, Orlando, Florida.
- Kim, C-W (1965) On the autonomy of the tensivity feature in stop classification (with special reference to Korean stops), *Word* 21: 339-356.
- Kim, C-W (1970) A theory of aspiration, *Phonetica*, 21: 107-116.
- Kim, D.-W., Hirose, H. and Niimi, S. (1992) A fiberoptic study of laryngeal gestures for Korean intervocalic stops, *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 26 : 13-28.
- Kim, D.-W., Niimi, S. and Hirose, H. (1993) Devoicing of vowel in Korean, *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 27 : 151-157.
- Kohler, K., J. (1979) Dimensions in the perception of fortis and lenis plosives, *Phonetica*, 36: 332-343.



- Kohler, K., J. (1982) F0 in the production of lenis and fortis plosives, *Phonetica*, 39: 199-218.
- Kohler, K., J. (1984) Phonetic explanation in phonology: The feature fortis/lenis, *Phonetica*, 41: 150-174.
- Kohler, K., J. (1990) Segmental reduction in connected speech in German: phonological facts and phonetic explanations, pp.69-92, in Hardcastle, W. J. and Marchal, A. (Eds.) *Speech Production and Speech Modelling*, Kluwer Academic Publishers, Dordrecht.
- Koizumi, T (1978) *Nihon-go no Seisho-hoo*, Taishuukan Shoten, Tokyo.
- Koizumi, T. (1989) Onsei to on'in, pp. 1-20, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (1)*, Meiji Shoin, Tokyo.
- Kori, S. (1989) Kyoochoo to Intonation, pp. 316-342, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (1)*, Meiji Shoin, Tokyo.
- Kubozono, H. (1989) The mora and syllable structure in Japanese: evidence from speech errors, *Language and Speech*, 32 (3), pp.249-278.
- Kubozono, H. (1991) *Nihon-go no Onsetsu-kyookai to Onsetsu-ryoo* (Syllable Boundary and Syllable Weight in Japanese) (in Japanese), A talk presented at a meeting of the Phonetic Association of Japan.
- Kubozono, H. (1995) The phonetic basis of phonological foot: evidence from Japanese, *The Proceedings of the Xiiith International Congress of Phonetic Sciences*, Vol. 4: pp. 372-375.
- Kubozono, H. (1995) Perceptual evidence for the mora in Japanese, in Connell, B., Arvaniti, A. and Watson, I. (Eds) *Papers in Laboratory Phonology IV*, Cambridge University Press.
- Kuehn, D. P. and Moll, K. L. (1976) A cineradiographic study of VC and CV articulatory velocities, *Journal of Phonetics*, 4: 303-320.
- Kuriyagawa, F. and Sawashima, M. (1984) Vowel duration in Japanese /hu'ku/ and /huku'/, *Annual Bulletin of Research Institute of Logopedics and Phoniatics*, 18: 83-90, Tokyo University.
- Kuriyagawa, F. and Sawashima, M. (1985) Vowel duration in Japanese /hu'ku/ and /huku'/(II), *Annual Bulletin of Research Institute of Logopedics and Phoniatics*, 19: 139-155, Tokyo University.
- Kuriyagawa, F. and Sawashima, M. (1986) Vowel duration in Japanese /tsu'ku/ and /tsuku'/, *Annual Bulletin of Research Institute of Logopedics and Phoniatics*, 20: 119-130, Tokyo University.
- Kuriyagawa, F. and Sawashima, M. (1987) Word accent and the duration of vowels in /C1u1C2u2/, *Annual Bulletin of Research Institute of Logopedics and Phoniatics*, 21: 41-52, Tokyo University.



- Kuriyagawa, F. and Sawashima, M. (1989) Word accent, devoicing and duration of vowels in Japanese, *Annual Bulletin of Research Institute of Logopedics and Phoniatics*, 23: 85-108, Tokyo University.
- Kuwabara, H. and Takeda, K (1988) Analysis and prediction of vowel-devocalization in isolated Japanese words, *ATR Technical Report TR-I-0030*, ATR Interpreting Telephony Research Laboratories, Kyoto.
- Ladefoged, P. (1982) *A Course in Phonetics (2nd)*, Harcourt Brace Jovanovich, New York.
- Laver, J. (1980) *The Phonetic Description of Voice Quality*, Cambridge University Press, Cambridge.
- Laver, J. (1994) *Principles of Phonetics*, Cambridge University Press, Cambridge.
- Lehiste, I. (1970) *Suprasegmentals*, MIT Press, Cambridge, Massachusetts.
- Lieberman, P. (1960) Some acoustic correlates of word stress in American English, *Journal of the Acoustical Society of America*, Vol. 32: 451.
- Lieberman, P. and Blumstein, S. E. (1988) *Speech Physiology, Speech Perception, and Acoustics Phonetics*, Cambridge University Press, Cambridge.
- Lindblom, B. (1963) Spectrographic study of vowel reduction, *Journal of the Acoustical Society of America*, Vol. 35, No. 11: 1773-81.
- Lindblom, B. (1983) Economy of speech gestures, pp. 217-245, in MacNeilage, P.F. (Ed.) *The Production of Speech*, Springer-Verlag, New York.
- Lovins, J. B. (1976) Pitch accent and vowel devoicing in Japanese : A preliminary study, *Annual Bulletin of Research Institute of Logopedics and Phoniatics* 10: 113-125.
- McCawley, J. (1977) Accent in Japanese, pp. 261-302, in Hyman, M. L. (Ed.) *Studies in Stress and Accent, Southern California Occasional Papers in Linguistics* No. 4, The department of Linguistics, University of Southern California, Los Angeles.
- Maekawa, K. (1989) Boin no musei-ka, pp. 135-153, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (1)*, Meiji Shoin, Tokyo.
- Maekawa, K. (1990a) Effects of speaking rate on the voicing variation in Japanese, *Journal of Electronic Information Communication Engineering*, , SP 89 - 148: 47-53.
- Maekawa, K. (1990b) Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese, *Proceedings of International Conference on Spoken Language Processing*, pp. 517-520, Kobe.
- Manuel, S. Y., Shattuck-Hufnagel, S., Huffman, M., Stevens, K., Carlson, R., and Hunnicutt S. (1992) Studies of vowel and consonant reduction, *The Proceedings of International Conference on Spoken Language Processing 92*, Banff.

- Mermelstein, P. (1973) Articulatory model for the study of speech production, *Journal of the Acoustical Society of America* 53: 1070-83.
- Miyatake, M. and Sagisaka, Y (1990) Prosodic characteristics and their control in Japanese speech with various speaking styles (in Japanese), *Journal of Electronic Information Communication Engineering D-II* Vol. J73-D-II, No. 12 : 1929-1935.
- Morin, Y. (1978) The status of mute 'e', *Studies in French Linguistics* 1 (2): 79-140.
- Morita, T. (1977) On'in no henshen (3), pp. 253-280, in Ohno, S. and Shibata, T. (Eds.) *Iwanami Kooza Nihon-go: Vol. 2 On'in*, Iwanami Shoten, Tokyo.
- Newton, B. (1972) *The Generative Interpretation of Dialect - A Study of Modern Greek Phonology*, Cambridge University Press, London.
- Nagano-Madsen, Y. (1994) Vowel devoicing rates in Japanese from a sentence corpus, *Working Papers of the Department of Linguistics*, Lund University, No. 42, pp. 119-129.
- Ogawa, K. (1981) A crosslinguistic study of voiceless vowels and some related phenomena, *Annual Reports of The Division of Languages*, International Christian University, Vol. 6, Tokyo.
- Ohso, M., (1973) A phonological study of some English loan words in Japanese, *Working Papers in Linguistics*, No. 14: 1-26, The Ohio State University, Ohio.
- Ohtaka, H. (1988) Nihon-go ni okeru boin musei-ka no ichi kousatsu (A brief note on Japanese vowel devoicing) (in Japanese), *a handout of the 6th Seminar on English Phonology*.
- Peterson, G. E. and Lehiste, I. (1960) Duration of syllable nuclei in English, *JASA*, Vol.32, No. 6.
- Pickett, J. M. (1980) *The Sounds of Speech Communication: A Primer of Acoustic Phonetics and Speech Perception*, PRO-ED, Austin, Texas.
- Pierrehumbert, J. B. and Beckman, M. E. (1988) *Japanese Tone Structure*, The MIT Press, Cambridge, Massachusetts.
- Port, R. F., Al-Ani, S. and Maeda, S. (1980) Temporal compensation and universal phonetics, *Phonetica* 37: 235-252.
- Port, R. F., Dalby, J. and O'Dell, M. (1987) Evidence for mora timing in Japanese, *Journal of the Acoustical Society of America*, Vol. 81, No. 5, May: 1574-1585.
- Poser, W. (1984) *The Phonetics and Phonology of Tone and Intonation in Japanese*, Ph. D dissertation, MIT.
- Poser, W. (1986) Japanese evidence bearing on the compensatory lengthening controversy, pp. 167-186, in Wetzels, L. and Sezer, E. (Eds.) *Studies in Compensatory Lengthening*, Foris Publications, Dordrecht.
- Poser, W. (1990) Evidence for foot structure in Japanese, *Languages* 66 (1): 78-105.



- Prince, A. and Smolensky, P. (1993) Optimality theory -- Constraint interaction in generative grammar, *Technical Report #2 of the Rutgers Center for Cognitive Science*, Rutgers University.
- Sagisaka, Y. and Tohkura, Y. (1984a) Phoneme duration control for speech synthesis by rule (in Japanese), *Journal of Electronic Information Communication Engineering*, 7, Vol. J67 - A No.7 : 629-636.
- Sagisaka, Y. and Sato, H. (1984b) Accentuation rules for Japanese Text-to-Speech conversion, *Review of the Electrical Communication Laboratories* Vol. 32 No. 2:188-199.
- Sagisaka, Y., Takeda, K., Abe, M., Katagiri, S., Umeda, T. and Kuwabara, H. (1990) A large-scale Japanese speech database, *Proceedings of International Conference on Spoken Language Processing*, pp.1089-1092, Kobe.
- Sakurai, S. (1985) Kyootsuu-go no hatsuon de chuui subeki kotogara, pp. 128-134, Appendix to NHK (Ed.), *NHK Accent Dictionary*, Tokyo.
- Sato, H. (1978) Temporal characteristics of spoken words in Japanese, *Journal of the Acoustical Society of America*, Vol. 64, Supplement No.1, Fall: S113-114.
- Sato, Y. (1993) The durations of syllable-final nasals and the mora hypothesis in Japanese, *Phonetica* 50: 44-67.
- Sawashima, M. (1969) Devoiced syllables in Japanese : A preliminary study by photoelectric glottography, *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, 3: 35-41.
- Sawashima, M. (1971) Devoicing of Vowels, *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 5: 7-13, Tokyo University.
- Schröder, M. and Schröder, H. (1987) Voiceless vowels in Toposa, *Afrikanistische Arbeitspapiere* 12: 17-26.
- Semiloff-Zelasko, H. (1973) Vowel reduction and loss in modern Hebrew fast speech, *Hebrew computational Linguistics* 7: 53-72.
- Shibata, T. (1962) On'in, reprinted in Shibata, T., Kitamura, H. and Kindaichi, H. (Eds.) (1980) *Nihon no Gengo-gaku : Vol. 2 On' in*, pp. 219-240, Taishuukan Shoten, Tokyo.
- Shibatani, M. (1990) *The Languages of Japan*, Cambridge University Press, Cambridge.
- Simada, Z. , Horiguchi, S., Niimi, S. and Hirose, H. (1991) Devoicing of Japanese /u/: an electromyographic study, *The Proceedings of the XIIth International Congress of Phonetic Science* Vol. 2: pp. 54-57.
- Smith, A. (1975) Syllable-structure and rhythm in Japanese, *Work in Progress*, 8: 1-13, Department of Linguistics, University of Edinburgh.
- Soli, S. (1981) Second formants in fricatives : Acoustic consequences of fricative-vowel coarticulation, *Journal of the Acoustical Society of America*, Vol. 70, No. 4 Oct: 976-984.



- Stevens, K. N. (1983) Design features of speech sound systems, pp. 247-261, in MacNeilage, P. F. (Ed.) *The Production of Speech*, Springer - Verlag, New York.
- Stevens, K. N. and House, A. S. (1963) Perturbation of vowel articulations by consonantal context: an acoustical study, *Journal of Speech and Hearing Research* 6: 111-128.
- Sugito, M. and Hirose, H. (1988) Production and perception of accented devoiced vowels in Japanese, *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 22 : 21-39.
- Sugito, M. (1989) Onsei ka haku ka, pp. 154-177, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (I)*, Meiji Shoin, Tokyo.
- Tabata, T. (1989) Nihongo no onsetsu koozoo ni tsuite: shuyoobu to shuushokubu, *Gengo Kenkyuu*, Vol. 95: 144-175.
- Takeda, K. and Kuwabara, H. (1987) Analysis and prediction of devocalizing phenomena (in Japanese), *Proceedings of Acoustic Society of Japan*, pp. 105-106.
- Takeda, K. , Sagisaka, Y. and Kuwabara, H. (1989) On sentence-level factors governing segmental duration in Japanese, *Journal of the Acoustic Society of America*, Vol. 86, No. 6 , Dec: 2081-2087.
- Thráinsson, H. (1978) On the phonology of Icelandic pre-aspiration, *Nordic Journal of Linguistics*, Vol.1, No. 1: 3-54.
- Thráinsson, H. (1980) Sonorant devoicing at Lake Mývatn: a change in progress, pp.355-364, in Hovdhaugen, E. (Ed.) *Proceedings of the Fourth International Conference of Nordic and General Linguistics in Oslo 1980*, universitetsforlaget, Oslo.
- Tranel, B. (1987a) French schwa and nonlinear phonology, *Linguistics* 25: 845-866.
- Tranel, B. (1987b) *The Sounds of French*, Cambridge University Press, Cambridge.
- Tsuchida, A. (1994) Fricative-vowel coarticulation in Japanese devoiced syllables: acoustic and perceptual evidence, *Working Papers of the Cornell Phonetics Laboratory*, Vol. 9, pp.183-221.
- Uemura, Y. (1989) Gojuu-onzu no onsei-gaku, pp. 41-63 in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (I)*, Meiji Shoin, Tokyo.
- Ueno, Z. (1977) Nihon-go no akusento, pp. 281-321, in Ohno, S. and Shibata, T. (Eds.) *Iwanami Kooza Nihon-go: Vol. 2 On'in*, Iwanami Shoten, Tokyo.
- Ueno, Z. (1989) Nihon-go no akusento, pp. 178-205, in Sugito, M. (Ed.) *Nihon-go no Onsei-On'in (I)*, Meiji Shoin, Tokyo.
- Vance, T. J. (1987) *An Introduction to Japanese Phonology*, State University of New York Press, Albany, N.Y. .

Weitzman, R., Sawashima, M., Hirose, H. and Ushijima, T. (1976) Devoiced and whispered vowels in Japanese, *Annual Bulletin of Research Institute of Logopedics and Phoniatics*, 10: 61-79, Tokyo University.

Wells, J., C. (1965) The phonological status of syllabic consonants in English RP, *Phonetica* 13: 110-113.

Wheeler, M. (1979) *Phonology of Catalan*, Basil Blackwell, Oxford.

Yeni-Komshian, G. H. and Soli, S. (1981) Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation, *J. Acoust. Soc. Am* 70(4) Oct. : 966-975.

Yoshida, N. and Sagisaka, Y. (1990) Factor analysis of vowel devoicing in Japanese (in Japanese), *ATR Technical Report TR-I-0159*, ATR Interpreting Telephony Research Laboratories, Kyoto.

Yoshida, S. (1990) A government-based analysis of the 'mora' in Japanese, *Phonology* 7: 331-351.

Yoshioka, H. (1981) Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese, *Phonetica*, 38: 236-251.

Yoshioka, H., Löfqvist, A. and Hirose, H. (1982) Laryngeal adjustments in Japanese voiceless sound production, *Journal of Phonetics*, 10: 1-10.

Zwicky, A. (1970) Auxiliary reduction in English, *Linguistic Inquiry* 1: 323-336.

Zwicky, A (1972) Note on a phonological hierarchy in English, pp. 275-301, in R. Stockwel and R. Macaulay (Eds.), *Linguistic Change and Generative Theory*, Indiana University Press, Bloomington.

# Appendix I (Chapter 3)

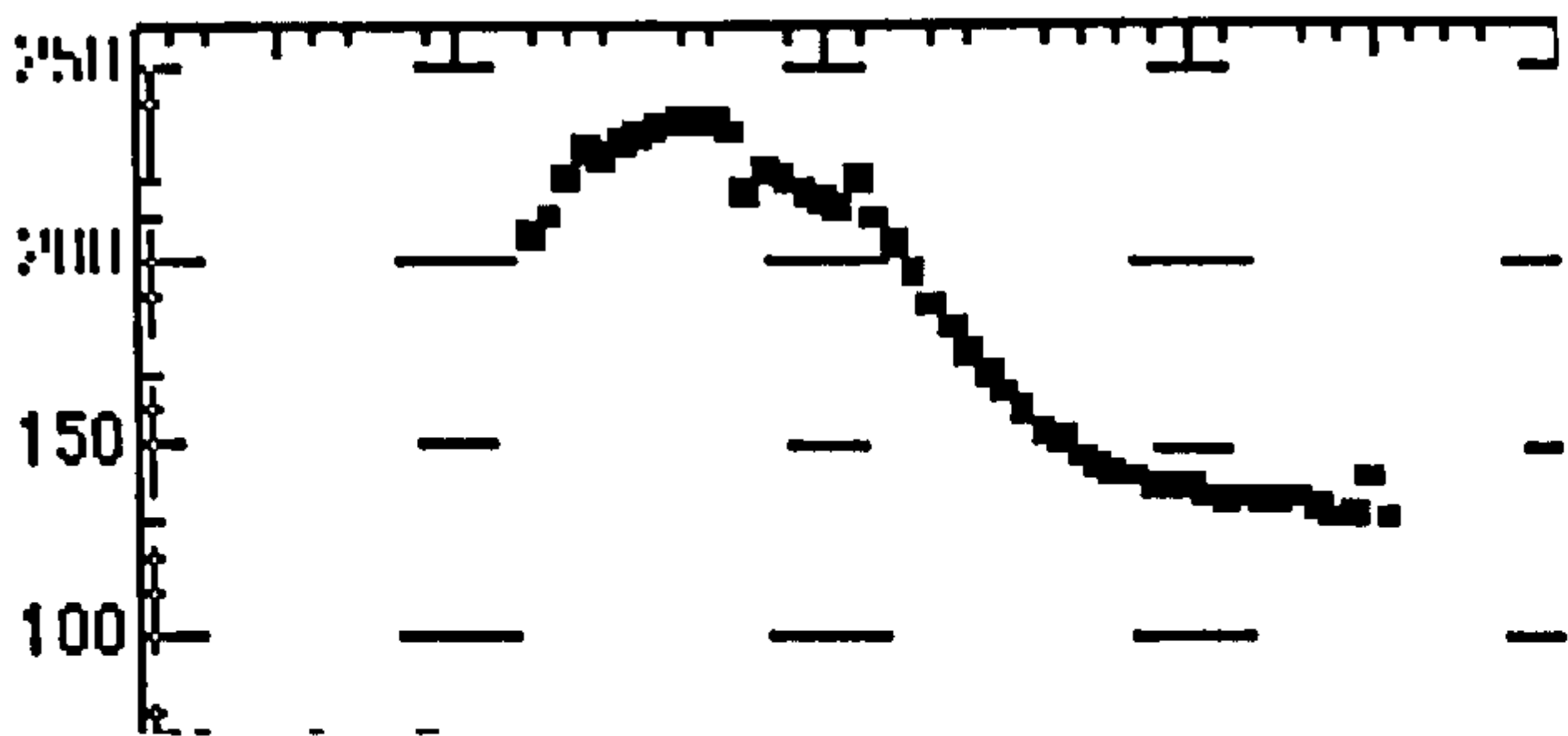
## Section 3.3

### Figures 1(a), 1(b) and 1(c)

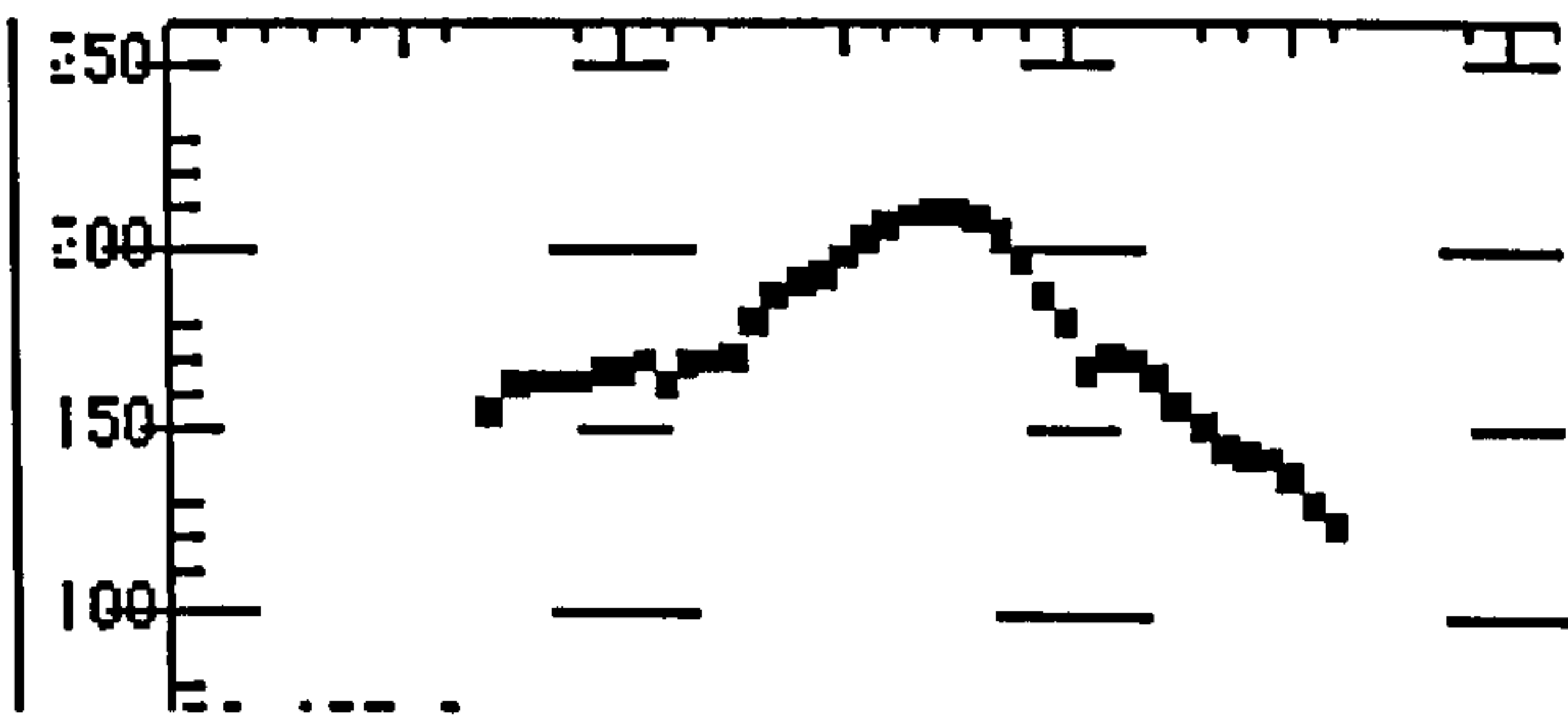
Three patterns of F0 contour of two-mora words followed by a subject marker -ga: (a) /a'niga/, (b) /ana'ga/, and (c) /anega/

1(a) The first mora is accented:

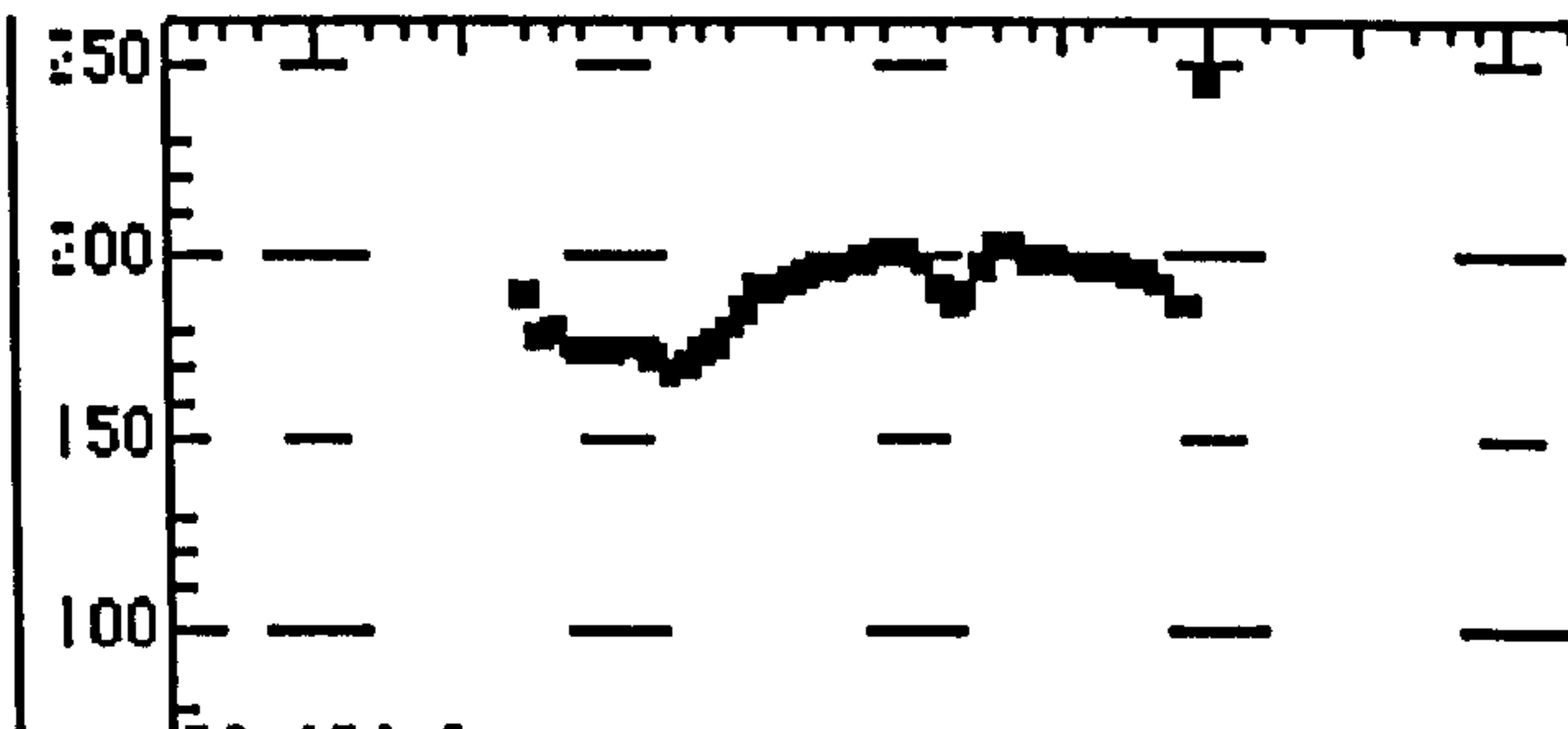
/a'niga/ 'my elder brother - Sub.marker'



1(b) The second mora is accented: /ana'ga/ 'hole- Sub.marker'



1 (c) No accent: /anega/ 'my elder sister - Sub.marker'





Section 3.4

Table 1 Allophonic alternation of obstruents across a morpheme boundary

(a)	[os-ɯ]	‘push (present)’	(b)	[mots-ɯ]	‘hold (present)’
	[os-anai]	‘push (negative)’		[mot-anai]	‘hold (negative)’
	[os-e]	‘push (imperative)’		[mot-e]	‘hold (imperative)’
	[oɕ-imas-ɯ]	‘push (polite)’		[motɕ-imas-ɯ]	‘hold (polite)’
	[oɕ-ita]	‘push (past)’		[mot-ta]	‘hold (past)’

Section 3.5.1

Table 2 Conservative Japanese consonant and vowel system  
(\* denotes the CV sequence is missing)

	/i/	/e/	/a/	/o/	/u/
(a) [s]	*	[se]	[sa]	[so]	[suɰ]
(b) [ɸ]	[ɸi]	*	[ɸa]	[ɸo]	[ɸuɰ]
(c) [t]	*	[te]	[ta]	[to]	*
(d) [z]	*	[ze]	[za]	[zo]	*
(e) [d]	*	[de]	[da]	[do]	*
(f) [ɸ]	*	*	*	*	[ɸuɰ]
(g) [ts]	*	*	*	*	[tsuɰ]
(h) [w]	*	*	[wa]	*	*

## Section 3.4

Table 3 Innovative Japanese consonant and vowel system

N.B. (F) = French, (G) = German, (I) = Italian, (L) = Latin, (R) = Russian

(a)	[si]: [si:di:]	'CD'	[i:si:]	'EC'
(b)	[çe]: [çe:kɯ]	'(milk) shake'	[çeri:]	'sherry'
(c)	[ti]: [pa:ti:]	'party'	[sɯpaget:i:]	'spaghetti'
	[tɯ]: [tɯ:rɯ]	'(F) Tours (place)'	[tɯ:]	'two'
(d)	[zi]: [zip:a:]	'zipper'		
	[zɯ]:			
(e)	[di]: [dizɯni: rando]	'Disneyland'	[disɯkɯ]	'disk'
	[dɯ]: [padɯdɯ:]	'(F) pas de deux'	[kisanadɯ:]	'Xanadu'
(f)	[fi]: [sɯfinkɯsɯ]	'(L) Sphinx'	[figɯa]	'figure skate'
	[fe]: [fɛŋçingɯ]	'fencing'	[feri:]	'ferryboat'
	[fa]: [fak:ɯsɯ]	'facsimile'	[sofa:]	'sofa'
	[fo]: [fo:kɯ]	'fork'	[forɯte]	'(I) forte'
(g)	[tsi]: [eritsin]	'(R) Yeltsin'	[sorɯdzenitsin]	'(R) Solzhenitsyn'
	[tse]: [konɰserɯŋ]	'(G) Konzern'	[fiɛɛntse]	'(I) Firenze'
	[tsa]: [tsa:]	'(R) Czar'	[pit:sa]	'pizza'
	[tso]: [kantso:ne]	'(I) canzone'	[paparət:so]	'(I) paparazzo'
(h)	[wi]: [wi:kɯ]	'week'	[harowi:ŋ]	'Halloween'
	[we]: [wedɛŋɯ]	'wedding'	[norɯwe:]	'Norway'
	[wo]: [wo:kɯman]	'walkman'	[wotsɯka]	'(R) vodka'

Other new phonetic sequences are: (a) [djɯ], e.g. [djɯet:o] 'duet', [pɯrodjɯ:sa:] 'producer', [djɯ:kɯ] 'duke', (b) [tjɯ], e.g. [tjɯ:ba], 'tuba', (c) [tçɛ], e.g. [tçesɯ] 'chess', [tçek:ɯ] (check) (Koizumi, 1987: 19).



Section 3.5.2

Table 4    List of the phonetic realisations of the moraic nasal by following sounds

(a) ~ (f) are examples of the moraic nasal before various consonants, and (g) is an example of the moraic nasal before vowels.

a.	/ka.N.-pa.N/	[kampan]	‘deck’
b.	/ka.N.-ba.N/	[kamban]	‘signboard’
c.	/ka.N.-ma.N/	[kam:an]	‘dullness’
d.	/ka.N.-ta.N/	[kantan]	‘easy’
e.	/ka.N.-da.N/	[kandan]	‘temperature’
f.	/ka.N.-se.N/	[kansen]	‘infection’
g.	/ka.N.-ze.N/	[kanzen]	‘perfection’
h.	/ka.N.-na.N/	[kan:an]	‘hardship’
i.	/ka.N.-si.N/	[kançin]	‘concern’
j.	/ka.N.-ke.N/	[kanken]	‘authority’
k.	/ka.N.-ge.N/	[kan:en]	‘resolution’
l.	/ka.N.re.N/	[kanren]	‘relation’
m.	/ka.N.-a.N/	[kanan]	‘thought’

(Examples (a) ~ (l) are taken from Koizumi, 1989: 14-15)

## Appendix II (Chapter 6)

(a) The reading text A (test words are underlined)

Mazu, Asahi-Kaigan to dookutsu no aida kara staatoshi, higashi ni susumimasu.

‘Asahi Beach’      ‘cave’

Akatsuchi-Yama to garasu-koojoo no aida o toori, danpukaa o oikoshite massugu

‘Mt. Akatsuchi’   ‘glassware factory’                      ‘damp truck’

ikimasu. Choofu-Kichi o nuke, kenritsu-kookoo no temae o magatte, heikiko no

‘Chofu Base’                      ‘state high school’                      ‘arsenal’

mawari o mawari, fuku-toshin kara deteiru Tsukushi-Tetsudoo o koemasu.

‘the hub of the city’                      ‘Tsukushi Railway’

Ryokuchi-kooen to hifuka no mawari o mawari, kutsu-koojoo to kutsushita-koojoo ni

‘forest park’   ‘dermatology clinic’                      ‘shoe factory’      ‘sock factory’

sotte kita e mukai, taki-fudoo o migi ni magatte, shigaku-kaikan to senshu-kyookai no

‘private school hall’   ‘athletic association’

aida o toori, Nachi-Keikoku e detara, Hitachi-kaigan ni sotte aruku to suizoku-kan ni

‘Nachi Gorge’                      ‘Hitachi Beach’                      ‘aquarium’

tsukimasu.

### (Translation)

First of all, start between Asahi Beach and the cave and go east. Go through between Mr. Akatsuchi and the Glass factory, overtake the damp truck and carry on. Go through the Chofu Base, turn just before the state high school and go around the arsenal, and pass the Tsukushi Railway coming from the outskirt of the city. Pass between the forest park and the dermatology clinic, and go north along the shoe factory and sock factory, turn right at the fudoo temple, then pass between the private school hall and the athletic association. When you reach Nachi Gorge, walk along Hitachi Beach, then you'll see the aquarium.

## (b) The reading text B (test words are underlined)

Kuni no desaki-kan no shoki-kan ni yoruto, Gifu-Chihoo to Aichi-Chihoo wa,  
 'local office of      'cabinet                      'Gifu District' 'Aichi District'  
 the government'      secretary'

Kinki-Chihoo to iu shoohi-chi o hikae, kasaihoochi-ki ya kirifuki-ki, tetsukabuto ya  
 'Kinki District' 'consumer belt'              'fire alarm'              'sprayer' 'steel helmet'

mashikaku no garasu-ki o tsukutteiru. Ippoo, Toohoku-Chihoo wa, bokuchiku no  
 'square'      'glassware'                      'Toohoku District' 'livestock farming'

mondai o kokufukushi, sangyoo-kai wa shoohi-keikaku ni motozuita takaku-keiei ni  
 'restoration'                      'consumer planning'              'diversification'

noridashi, saikin koushu-kei haishi de wadaini natta Kasupi-Kai engan no kuniguni e  
 'death by hanging'                      'the Caspian Sea'

no shaapu-penshiru no jushutsu-kichi to natta. Mata, kapuchiino no ryuukoo de,  
 'propelling pencil' 'exporting centre'              'cappuccino'

Itaria e no kakuyasu-tsuaa ga ninki dearu.  
 'cheap package holiday'

## (Translation)

According to the cabinet secretary of the government office in the town, Gifu and Aichi Districts produce firealarms, sprayers, steel helmets and square glassware, having a consumer belt such as Kinki District. On the other hand, Toohoku District has overcome the problem of livestock farming, diversified its industry based on a consumer planning, and has become the exporting centre of propelling pencils to countries around the Caspian Sea, which were the top of the headlines for a while as they abolished the death by hanging law. And cheap package tours to Italy have been very popular because of the popularity of cappuccino.